A photograph of three deer standing in a grassy field. The deer on the left and middle are brown with white patches on their hindquarters, while the deer on the right is entirely white. They are standing in front of a background of bare trees with a warm, golden light, suggesting a sunset or sunrise. The text is overlaid in white, bold, sans-serif font.

# Unit 6

# Molecular Biology

## Chapter 14: Gene Expression: From Gene to Protein

# Overview: The Flow of Genetic Information

---

- The information content of genes is in the form of specific sequences of nucleotides in DNA
- The DNA inherited by an organism leads to specific traits by dictating the synthesis of proteins
- Proteins are the links between genotype and phenotype
- **Gene expression** is the process by which DNA directs protein synthesis
  - Includes two stages:
    - Transcription
    - Translation

## Concept 14.1: Genes specify proteins via transcription and translation

---

- How was the fundamental relationship between genes and proteins discovered?
- Garrod first suggested that genes dictate phenotypes through enzymes that catalyze specific chemical reactions
  - He thought symptoms of an inherited disease reflect an inability to synthesize a certain enzyme
  - Cells synthesize and degrade molecules in a series of steps, a metabolic pathway
- Beadle and Tatum developed the *one gene-one enzyme hypothesis*
  - The function of a gene is to dictate the production of a specific enzyme

# *The Products of Gene Expression: A Developing Story*

---

- Some proteins are not enzymes, so researchers later revised the one gene—one enzyme hypothesis:
  - One gene—one protein
- Many proteins are composed of several polypeptides, each of which has its own gene
- Therefore, Beadle and Tatum's hypothesis is now restated as the one gene—one polypeptide hypothesis
  - It is common to refer to gene products as proteins rather than polypeptides

# Basic Principles of Transcription and Translation

---

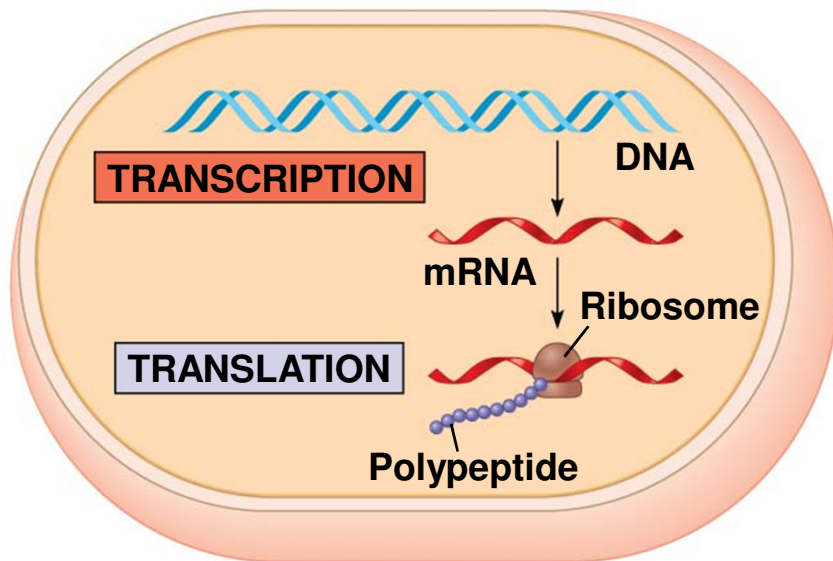
- Genes provide the instructions for making specific proteins
- RNA is the bridge between DNA and protein synthesis
- RNA is chemically similar to DNA, but RNA
  - Has a ribose sugar instead of deoxyribose
  - Has the base uracil (U) rather than thymine (T)
  - Is usually single-stranded
- Getting from DNA to protein requires two stages:
  - Transcription
  - Translation

- 
- **Transcription** is the synthesis of RNA using information in DNA
    - Produces **messenger RNA (mRNA)**
  - **Translation** is the synthesis of a polypeptide, using information in the mRNA
    - **Ribosomes** are the sites of translation

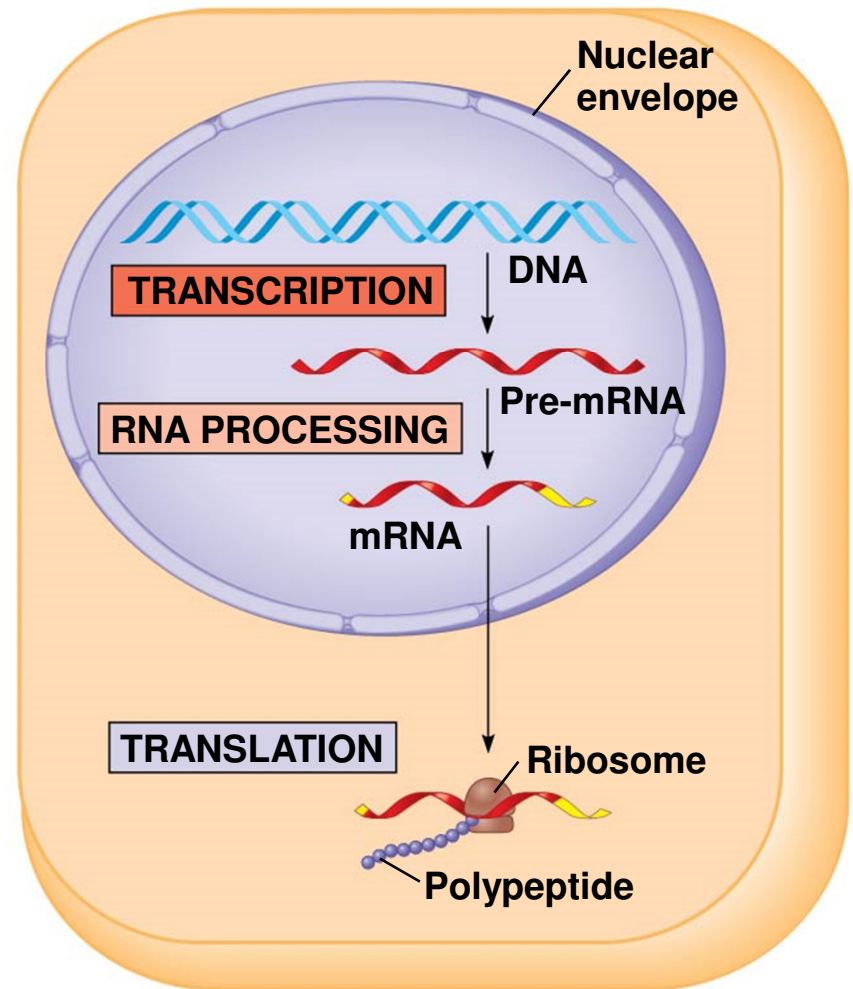
- 
- In prokaryotes, translation of mRNA can begin before transcription has finished
    - Due to lack of compartmentalization
  - In eukaryotes, the nuclear envelope separates transcription from translation
    - Transcription occurs in nucleus
    - Translation occurs in cytoplasm at ribosomes
  - Eukaryotic RNA transcripts are modified through RNA processing to yield the finished mRNA
  - Eukaryotic mRNA must be transported out of the nucleus to be translated



Figure 14.4



(a) Bacterial cell



(b) Eukaryotic cell



- 
- A **primary transcript** is the initial RNA transcript from any gene prior to processing
  - The *central dogma* is the concept that cells are governed by a cellular chain of command



# The Genetic Code

---

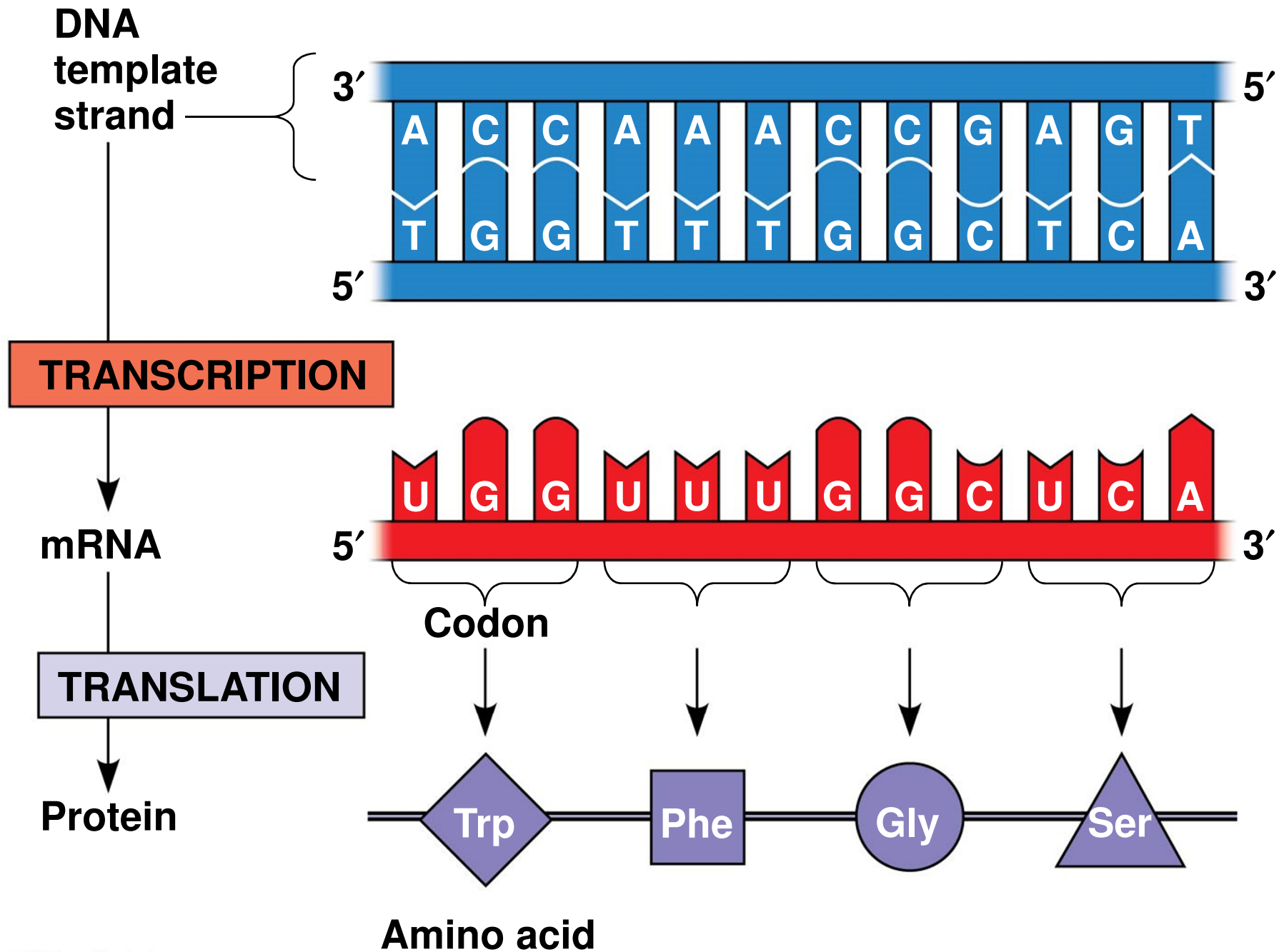
- How are the instructions for assembling amino acids into proteins encoded into DNA?
- There are 20 amino acids, but there are only four nucleotide bases in DNA
  - Thus each base cannot specify one amino acid
- How many nucleotides correspond to an amino acid?

# *Codons: Triplets of Nucleotides*

---

- The flow of information from gene to protein is based on a **triplet code**
  - A series of nonoverlapping, three-nucleotide words
- The words of a gene are transcribed into complementary nonoverlapping three-nucleotide words of mRNA
  - mRNA base triplets are called **codons**
- These words are then translated into a chain of amino acids, forming a polypeptide

Figure 14.5



- 
- During transcription, one of the two DNA strands, called the **template strand**, provides a template for ordering the sequence of complementary nucleotides in an RNA transcript
    - The template strand is always the same strand for any given gene
  - The mRNA molecule is complementary to the DNA template strand according to the base-pairing rules
    - But remember, RNA does NOT have thymine
    - So in RNA, uracil is brought in to pair with DNA's adenine
  - mRNA is synthesized in the 5' to 3' direction

- 
- During translation, the mRNA base triplets (codons) are read in the 5' to 3' direction
  - Each codon specifies the amino acid (one of 20) to be placed at the corresponding position along a polypeptide



# *Cracking the Code*

---

- All 64 codons were deciphered by the mid-1960s
  - The first codon (UUU) was deciphered by Marshall Nirenberg
- Of the 64 triplets, 61 code for amino acids
  - 3 triplets are “stop” signals to end translation
- The codon AUG has a dual function
  - Codes for amino acid methionine (Met)
  - Functions as a “start” signal

- 
- The genetic code is redundant
    - More than one codon may specify a particular amino acid
  - But it is not ambiguous
    - No codon specifies more than one amino acid
  - Codons must be read in the correct **reading frame** (correct groupings) in order for the specified polypeptide to be produced
    - Codons are read one at a time in a nonoverlapping fashion

Figure 14.6

		Second mRNA base					
		U	C	A	G		
First mRNA base (5' end of codon)	U	UUU	UCU	UAU	UGU	U C A G	Third mRNA base (3' end of codon)
		UUC	UCC	UAC	UGC		
		UUA	UCA	UAA Stop	UGA Stop		
		UUG	UCG	UAG Stop	UGG Trp		
	C	CUU	CCU	CAU	CGU	U C A G	
		CUC	CCC	CAC	CGC		
		CUA	CCA	CAA	CGA		
		CUG	CCG	CAG	CGG		
	A	AUU	ACU	AAU	AGU	U C A G	
		AUC	ACC	AAC	AGC		
		AUA	ACA	AAA	AGA		
		AUG Met or start	ACG	AAG	AGG		
	G	GUU	GCU	GAU	GGU	U C A G	
		GUC	GCC	GAC	GGC		
		GUA	GCA	GAA	GGA		
		GUG	GCG	GAG	GGG		

# *Evolution of the Genetic Code*

---

- The genetic code is nearly universal
  - Shared by the simplest bacteria and the most complex animals
- Genes can be transcribed and translated after being transplanted from one species to another

**(a) Tobacco plant expressing a firefly gene**



**(b) Pig expressing a jellyfish gene**



## Concept 14.2: Transcription is the DNA-directed synthesis of RNA: *a closer look*

---

- Transcription is the first stage of gene expression
- RNA synthesis is catalyzed by **RNA polymerase**
  - Pries the DNA strands apart and joins together the RNA nucleotides
- RNA polymerases assemble polynucleotides in the 5' to 3' direction
- Unlike DNA polymerases, RNA polymerases can start a chain without a primer

# Molecular Components of Transcription

---

- The DNA sequence where RNA polymerase attaches is called the **promoter**
- In bacteria, the sequence signaling the end of transcription is called the **terminator**
- The stretch of DNA that is transcribed is called a **transcription unit**



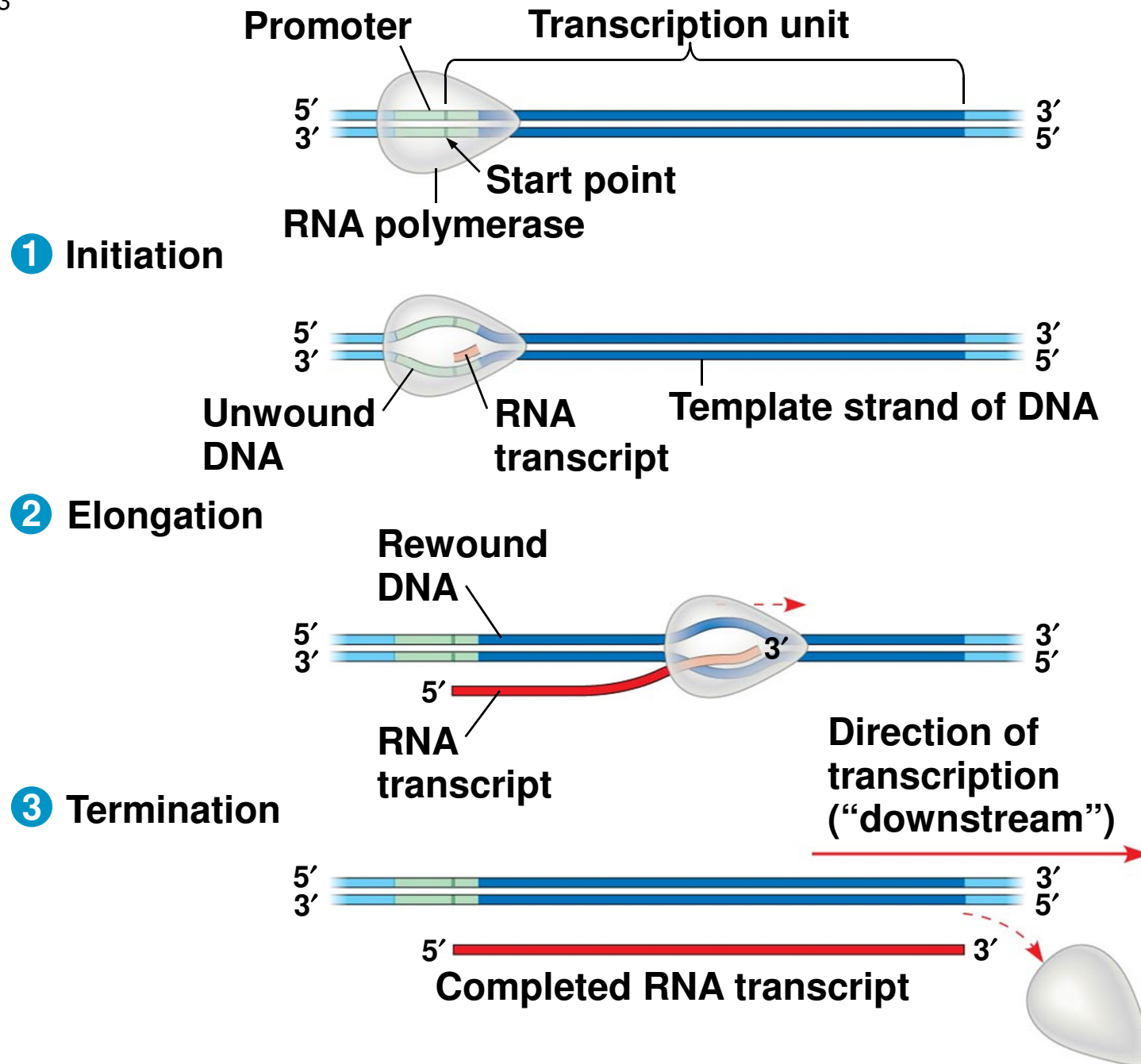
# Synthesis of an RNA Transcript

---

- The three stages of transcription
  1. Initiation
  2. Elongation
  3. Termination

**Animation: Transcription Introduction**  
Right click slide / Select Play

Figure 14.8-3

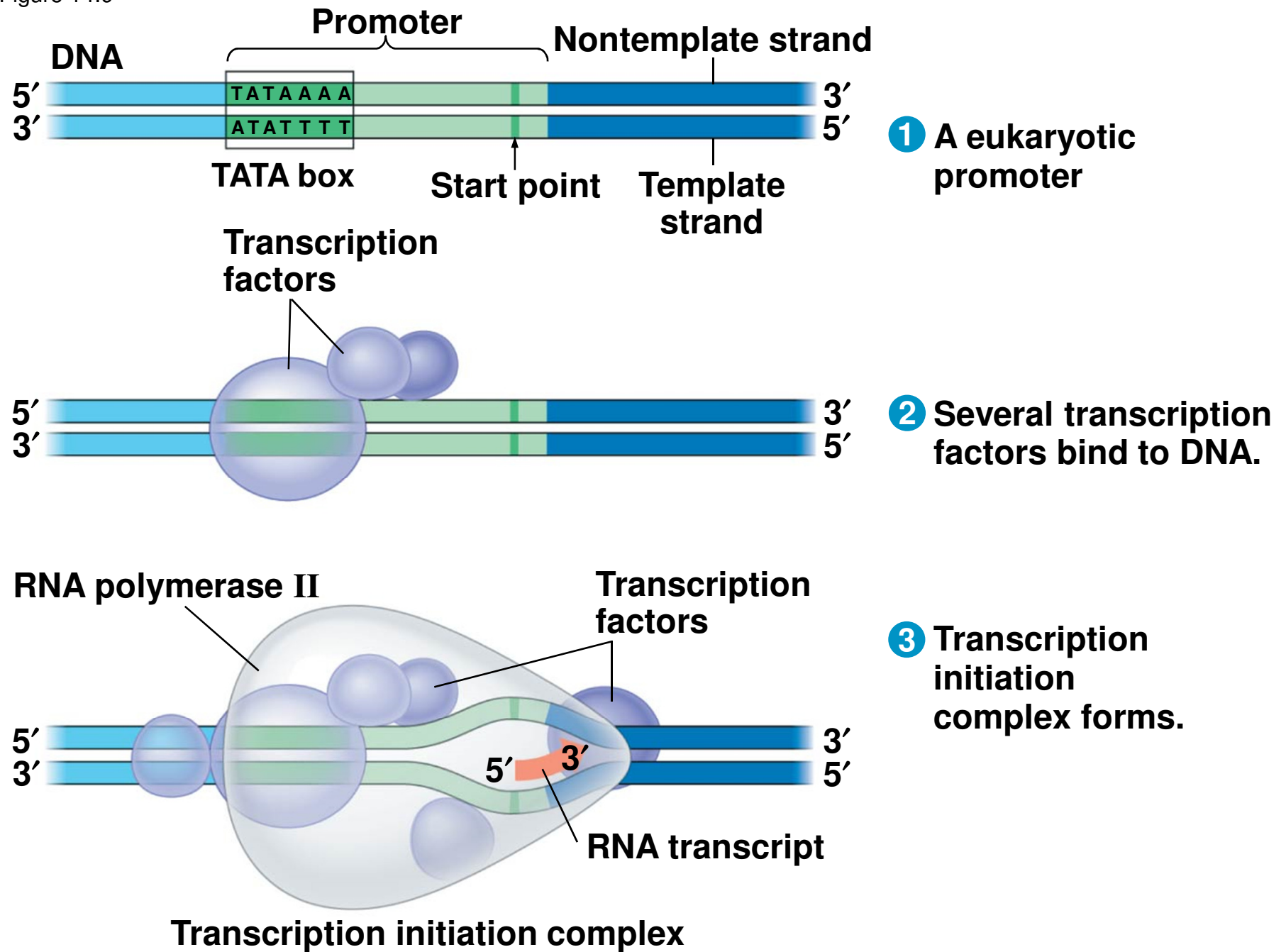


# *RNA Polymerase Binding and Initiation of Transcription*

---

- Promoters signal the transcriptional **start point** and usually extend several dozen nucleotide pairs upstream of the start point
- In bacteria, the RNA polymerase itself specifically recognizes and binds to the promoter
- In eukaryotes, **transcription factors** mediate the binding of RNA polymerase and the initiation of transcription
- The completed assembly of transcription factors and RNA polymerase II bound to a promoter is called a **transcription initiation complex**
- A promoter called a **TATA box** is crucial in forming the initiation complex in eukaryotes

Figure 14.9



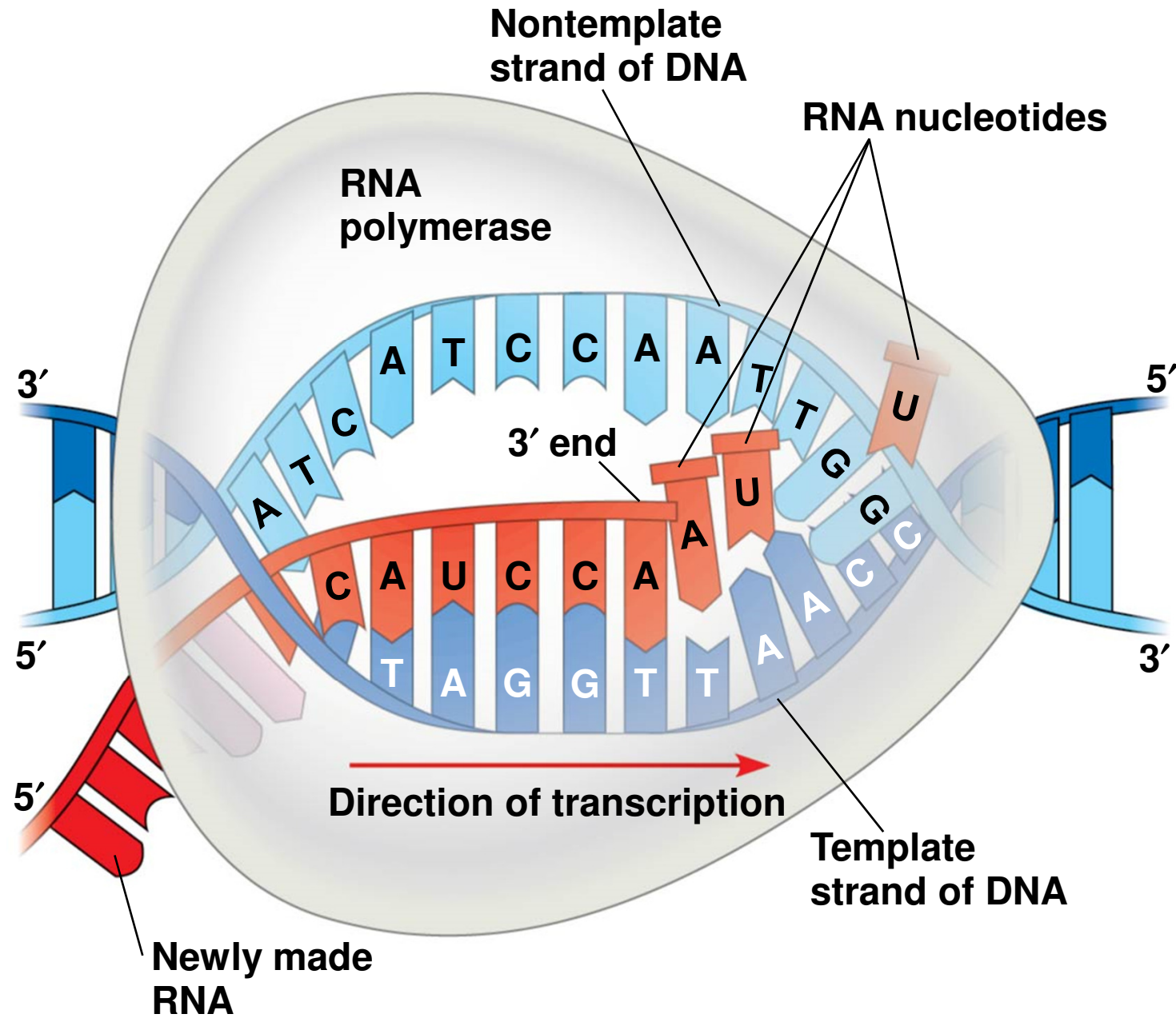
## *Elongation of the RNA Strand*

---

- As RNA polymerase moves along the DNA, it untwists the double helix
  - Adds nucleotides to the 3' end of the growing RNA molecule (*complementary* to DNA bases)
    - DNA has G; RNA pairs C
    - DNA has C; RNA pairs G
    - DNA has T; RNA pairs A
    - *DNA has A; RNA pairs U*
- A gene can be transcribed simultaneously by several RNA polymerases



Figure 14.10



# *Termination of Transcription*

---

- The mechanisms of termination are different in bacteria and eukaryotes
  - In bacteria
    - The polymerase stops transcription at the end of the terminator
    - mRNA can be translated without further modification
  - In eukaryotes
    - RNA polymerase II transcribes the polyadenylation signal sequence
    - The RNA transcript is released 10–35 nucleotides past this polyadenylation sequence

## Concept 14.3: Eukaryotic cells modify RNA after transcription

---

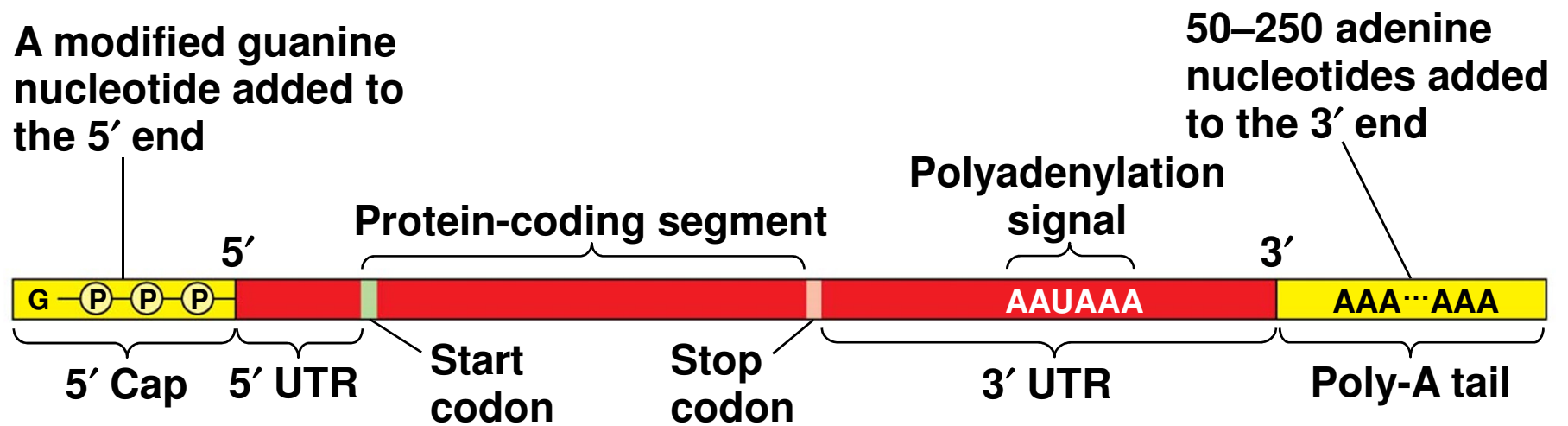
- Enzymes in the eukaryotic nucleus modify pre-mRNA (**RNA processing**) before the genetic messages are dispatched to the cytoplasm
- During RNA processing, both ends of the primary transcript are altered
- Also, usually some interior parts of the molecule are cut out and the other parts spliced together

# Alteration of mRNA Ends

---

- Each end of a pre-mRNA molecule is modified in a particular way
  - The 5' end receives a modified G nucleotide **5' cap**
  - The 3' end gets a **poly-A tail**
- These modifications share several functions
  - Facilitating the export of mRNA to the cytoplasm
  - Protecting mRNA from hydrolytic enzymes
  - Helping ribosomes attach to the 5' end

Figure 14.11

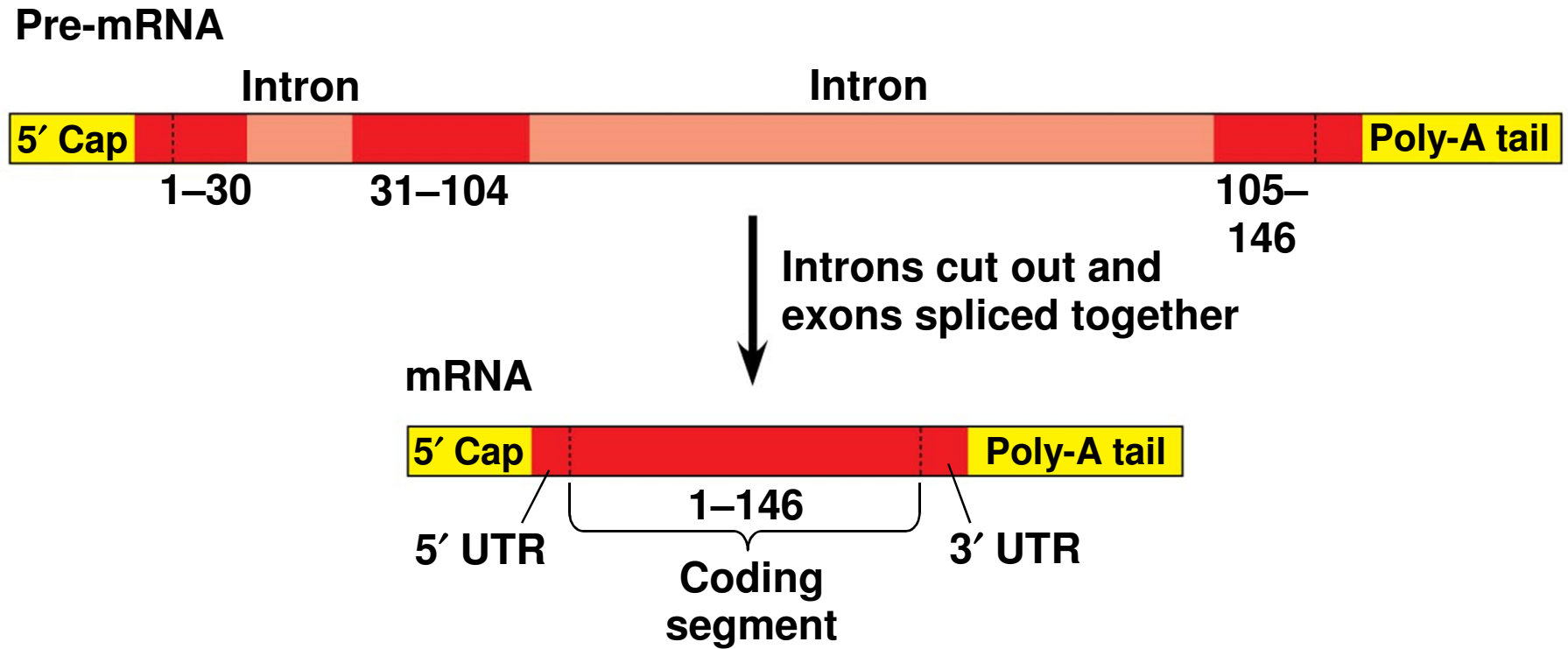


# Split Genes and RNA Splicing

---

- Most eukaryotic mRNAs have long noncoding stretches of nucleotides that lie between coding regions
- The noncoding regions are called intervening sequences, or **introns**
- The other regions are called **exons** and are usually translated into amino acid sequences
- **RNA splicing** removes introns and joins exons, creating an mRNA molecule with a continuous coding sequence

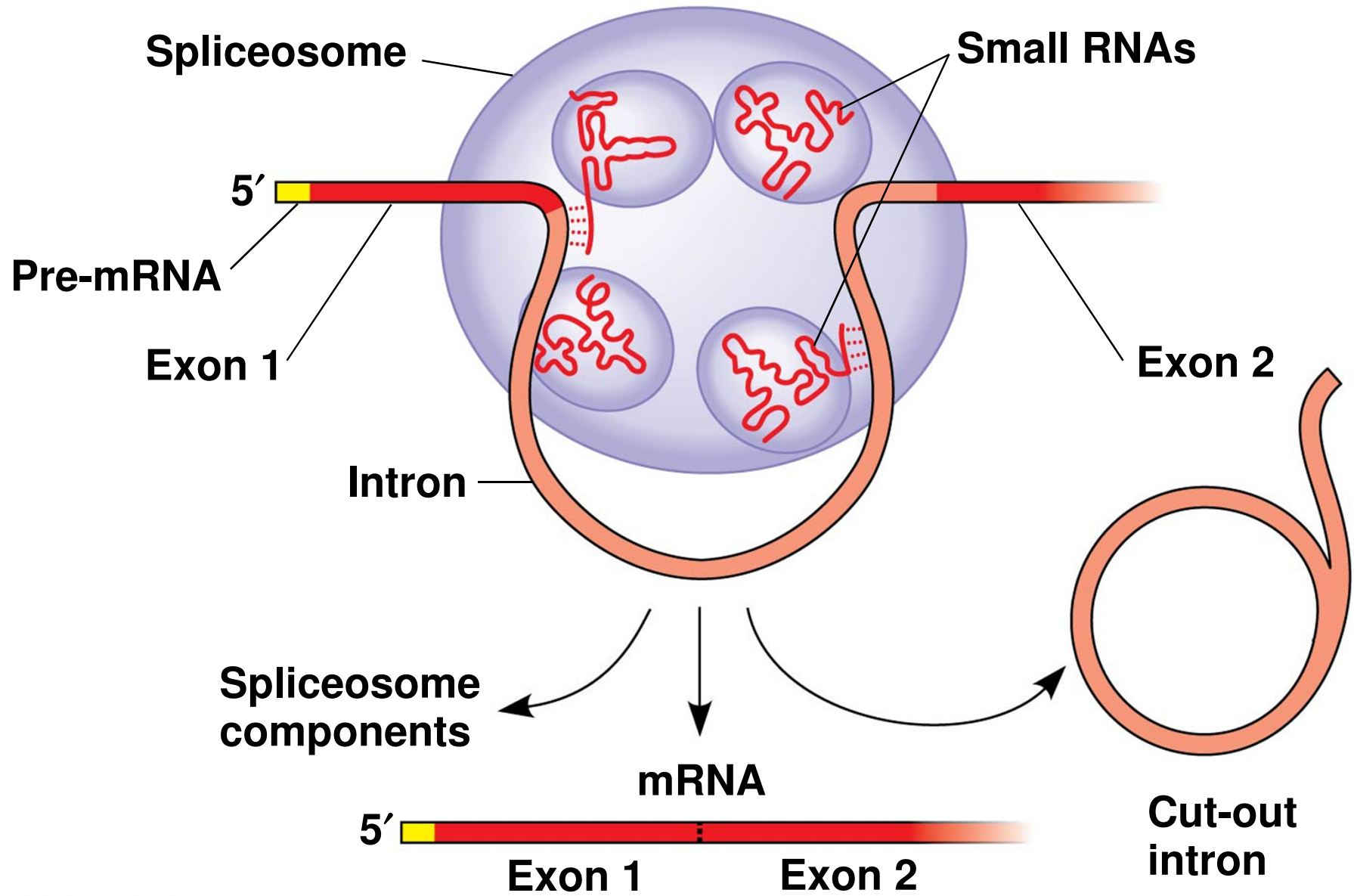
Figure 14.12



- 
- Many genes can give rise to two or more different polypeptides
    - Depending on which segments are used as exons
    - This process is called **alternative RNA splicing**
      - Because of this, the organism's number of different protein products can be much greater than it's number of genes
  - RNA splicing is carried out by **spliceosomes**
    - Complex of proteins and small RNAs that remove introns



Figure 14.13



# *Ribozymes*

---

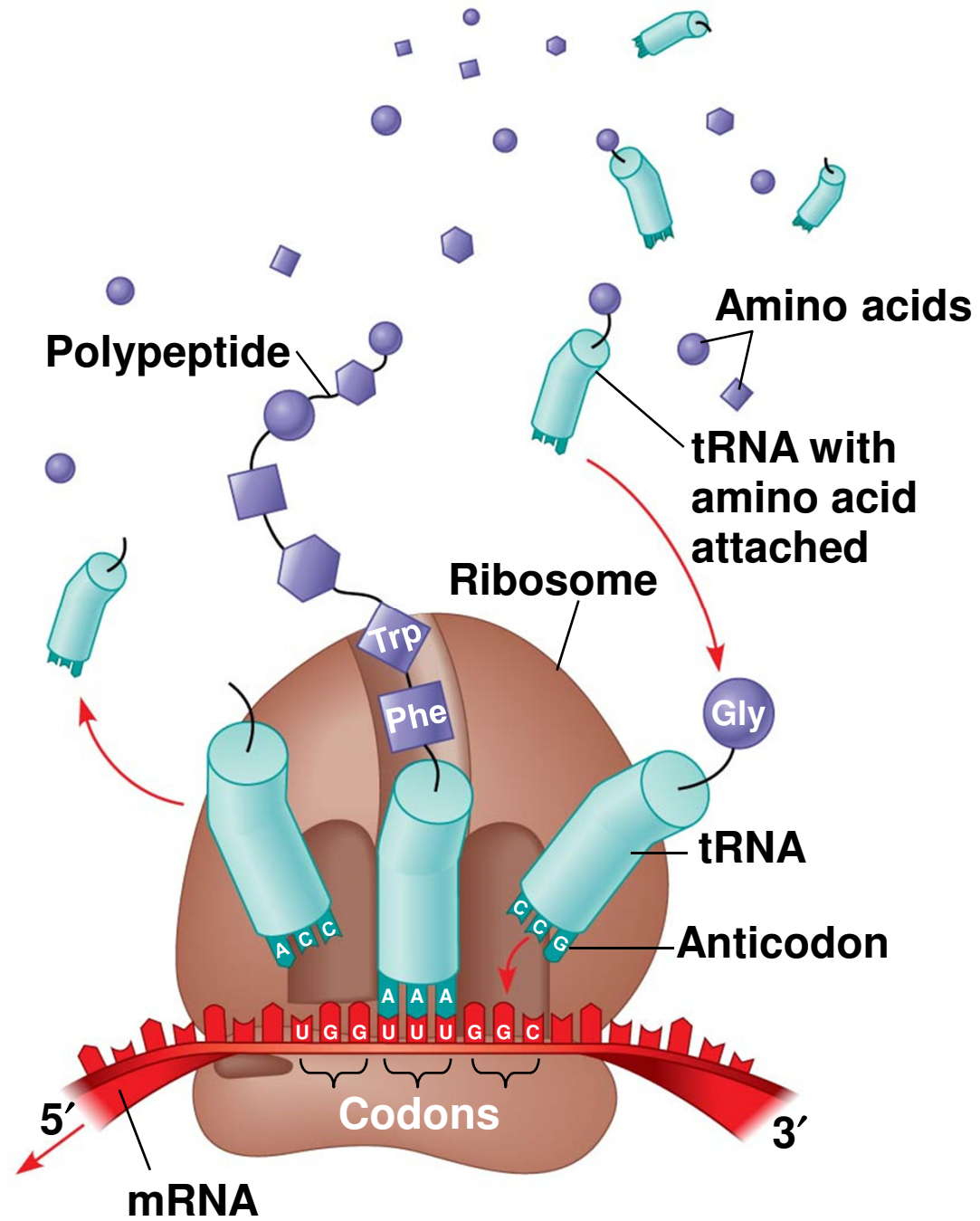
- **Ribozymes** are RNA molecules that function as enzymes
- RNA splicing can occur without proteins, or even additional RNA molecules
  - The introns can catalyze their own splicing
- 3 properties of RNA enable some RNA molecules to function as enzymes
  1. RNA is single stranded
    - May base-pair with complementary region elsewhere in the same molecule
  2. Some RNA bases contain functional groups that may participate in catalysis
  3. RNA can hydrogen-bond with other nucleic acids
    - Adds specificity to its catalytic activity

## Concept 14.4: Translation is the RNA-directed synthesis of a polypeptide: *a closer look*

---

- Genetic information flows from mRNA to protein through the process of translation
- A cell translates an mRNA message into protein with the help of **transfer RNA (tRNA)**
  - tRNAs transfer amino acids to the growing polypeptide in a ribosome

Figure 14.14



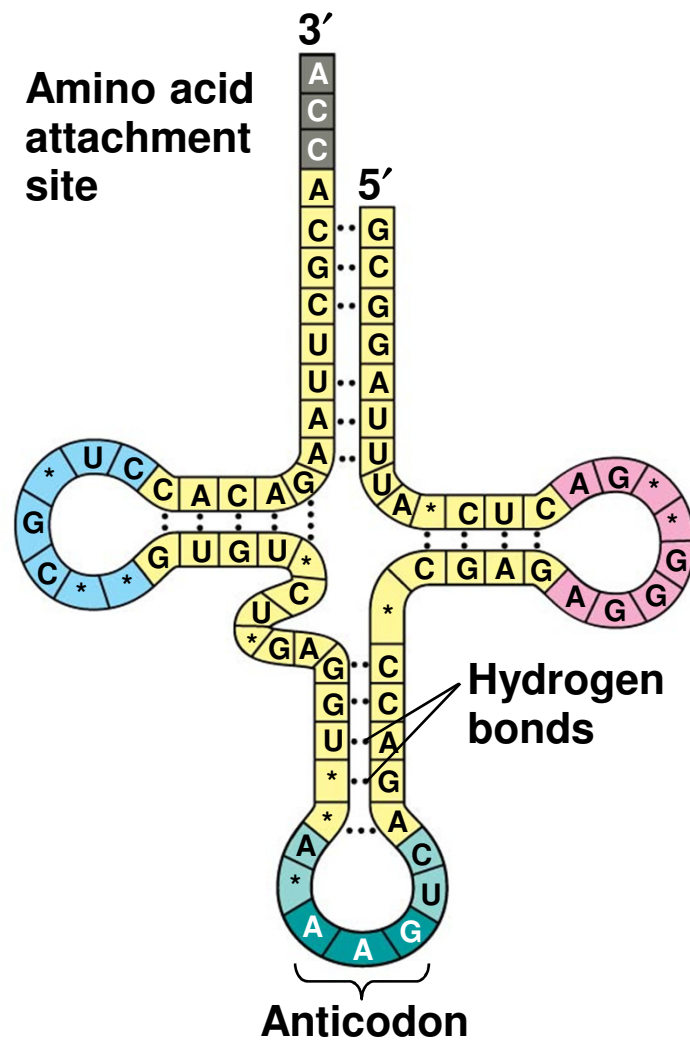
# *The Structure and Function of Transfer RNA*

---

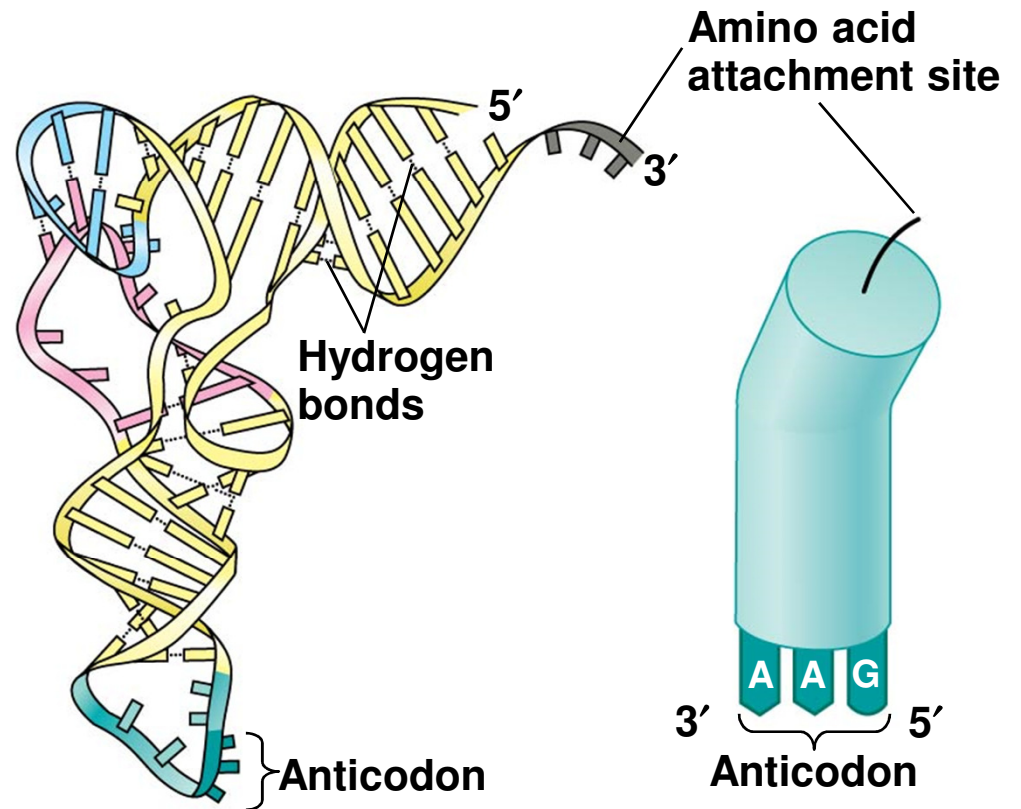
- Each tRNA can translate a particular mRNA codon into a given amino acid
  - The tRNA contains an amino acid at one end
  - At the other end has a nucleotide triplet that can base-pair with the complementary codon on mRNA

- 
- A tRNA molecule consists of a single RNA strand that is only about 80 nucleotides long
  - tRNA molecules can base-pair with themselves
  - Flattened into one plane, a tRNA molecule looks like a cloverleaf
  - In three dimensions, tRNA is roughly L-shaped, where one end of the L contains the **anticodon**
    - Nucleotide triplet that base-pairs with an mRNA codon
  - tRNA is a translator
    - Reads a nucleic acid word (mRNA codon)
    - Interprets it as a protein word (amino acid)

Figure 14.15



(a) Two-dimensional structure



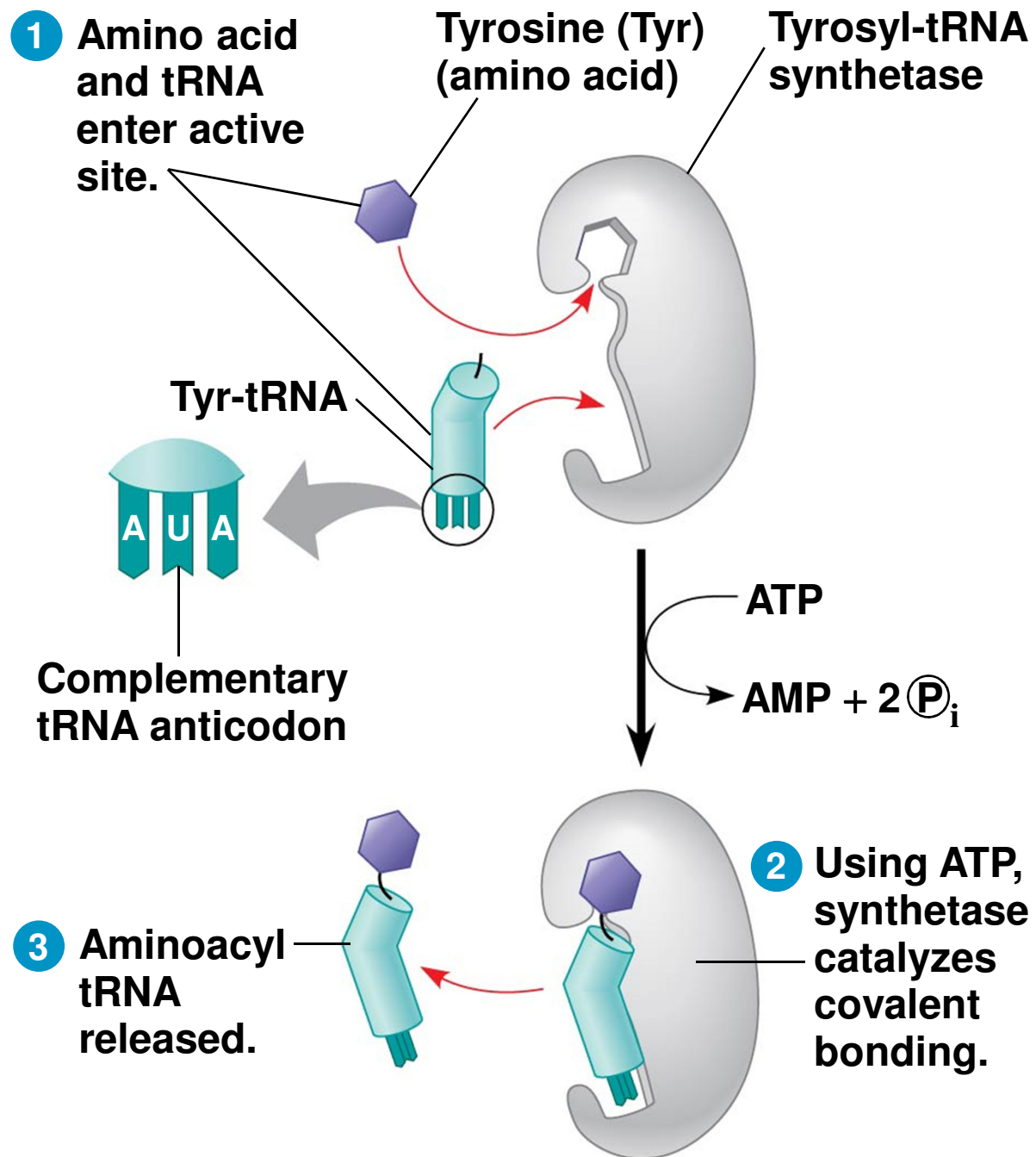
(b) Three-dimensional structure

(c) Symbol used in this book

- 
- Accurate translation requires two steps
    1. A correct match between a tRNA and an amino acid, done by the enzyme **aminoacyl-tRNA synthetase**
      - Active site of each aminoacyl-tRNA synthetase fits only a specific combination of amino acid and tRNA
    2. A correct match between the tRNA anticodon and an mRNA codon
  - Flexible pairing at the third base of a codon is called **wobble** and allows some tRNAs to bind to more than one codon



Figure 14.16-3



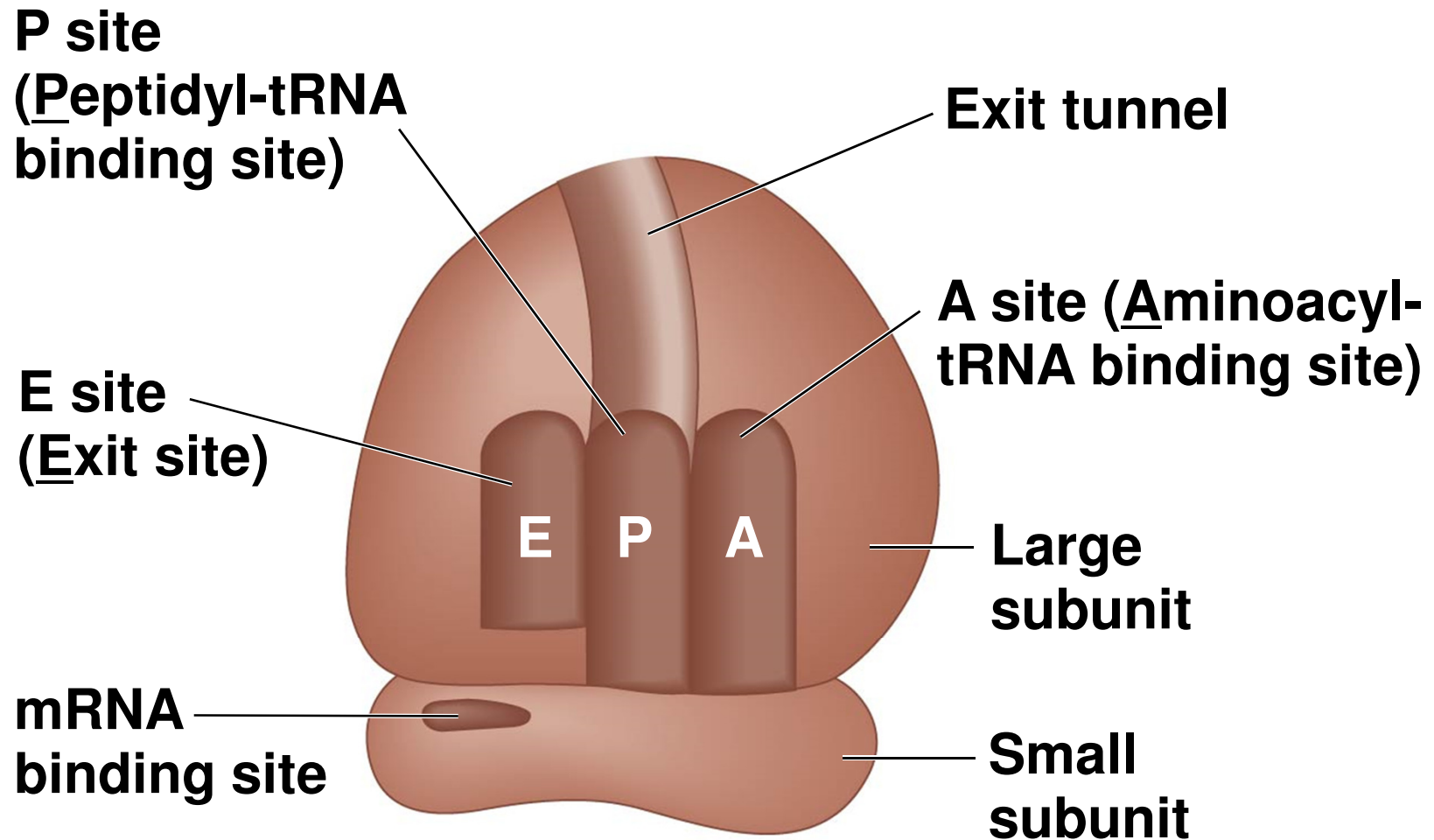
# *Ribosomes*

---

- Ribosomes facilitate specific coupling of tRNA anticodons with mRNA codons during protein synthesis
- The large and small ribosomal are made of proteins and **ribosomal RNAs (rRNAs)**
  - Made in nucleolus of eukaryotes
  - rRNA is the most abundant type of cellular RNA
- In bacterial and eukaryotic ribosomes the large and small subunits join to form a ribosome only when attached to an mRNA molecule

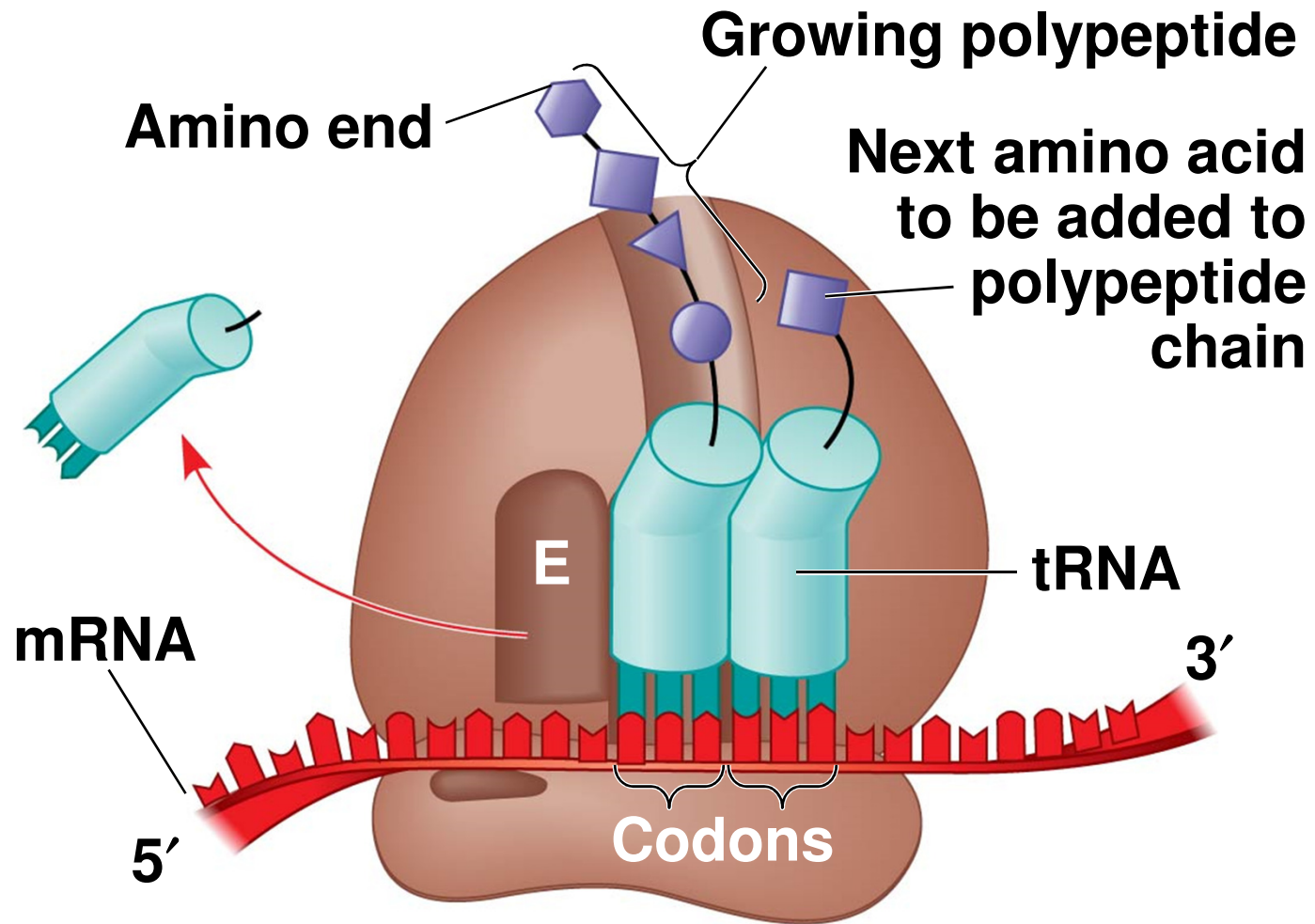
- 
- A ribosome has three binding sites for tRNA
    - The **P site** holds the tRNA that carries the growing polypeptide chain
    - The **A site** accepts the tRNA that carries the next amino acid to be added to the chain
    - The **E site** is the exit site, where discharged tRNAs leave the ribosome

Figure 14.17b



**(b) Schematic model showing binding sites**

Figure 14.17c

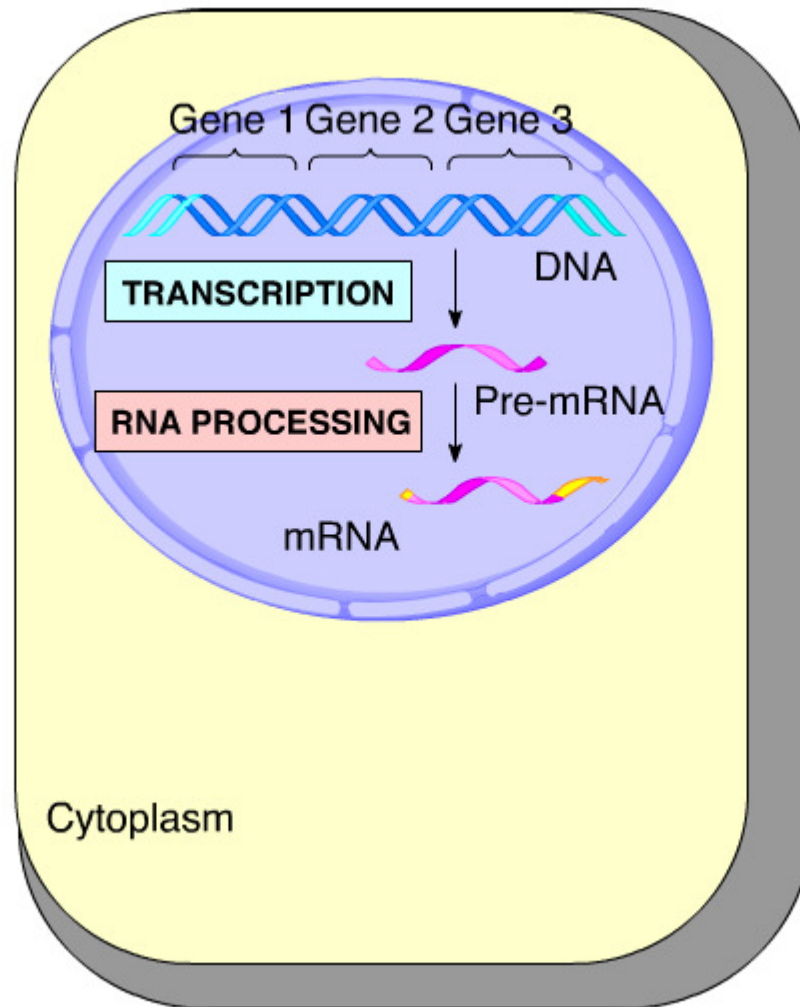


(c) Schematic model with mRNA and tRNA

# Building a Polypeptide

---

- The three stages of translation
  1. Initiation
  2. Elongation
  3. Termination
- All three stages require protein “factors” that aid in the translation process
- Energy is also required for certain aspects
  - Provided by the hydrolysis of guanosine triphosphate (GTP)



**Animation: Translation Introduction**  
Right click slide / Select play

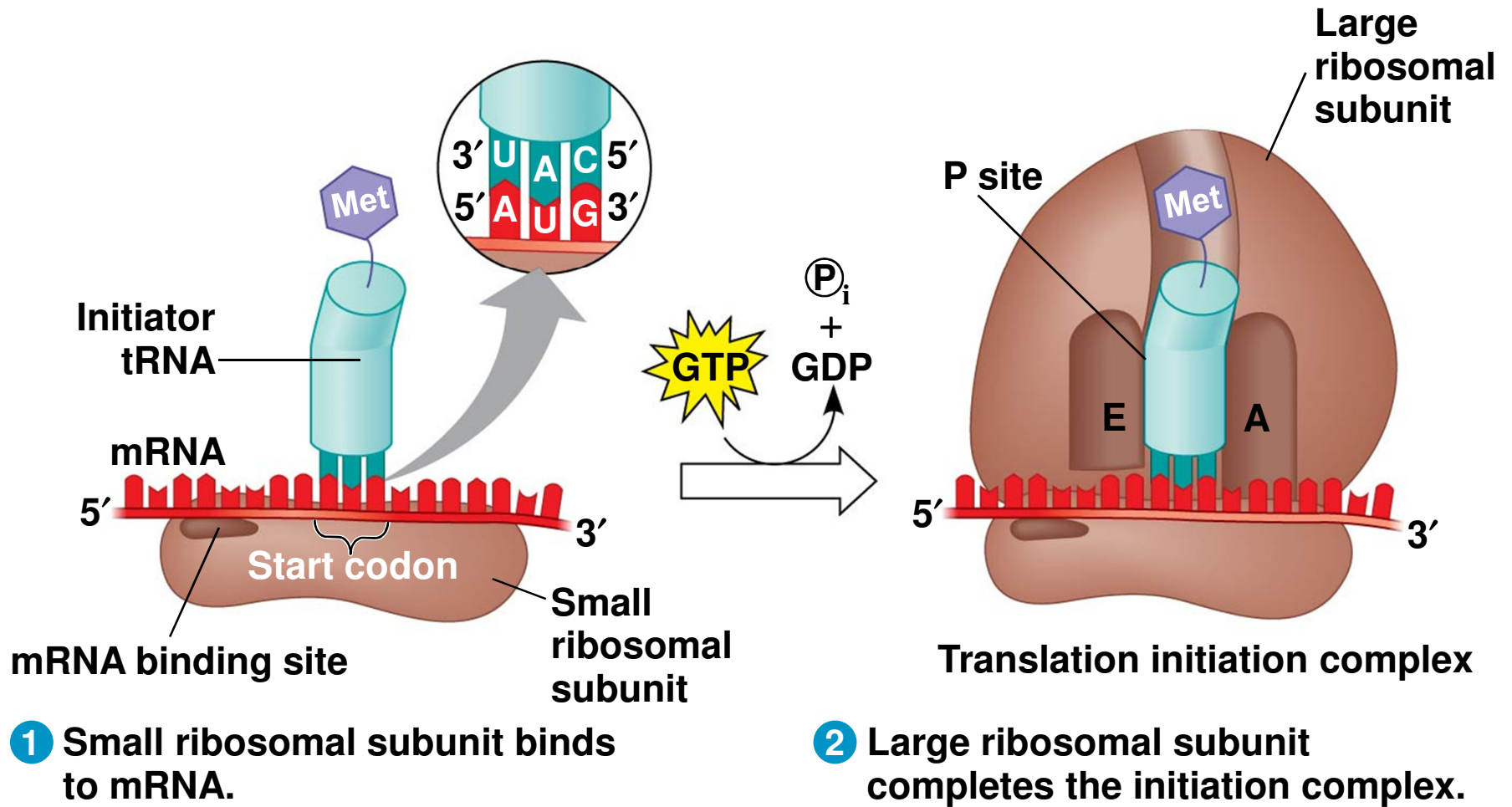
# ***Ribosome Association and Initiation of Translation***

---

- The initiation stage of translation brings together mRNA, a tRNA with the first amino acid, and the two ribosomal subunits
- A small ribosomal subunit binds with mRNA and a special initiator tRNA
- Then the small subunit moves along the mRNA until it reaches the start codon (AUG)
  - Establishes the codon reading frame for the mRNA
- The addition of the large ribosomal subunit is last and completes the formation of the *translation initiation complex*
- Proteins called *initiation factors* bring all these components together



Figure 14.18

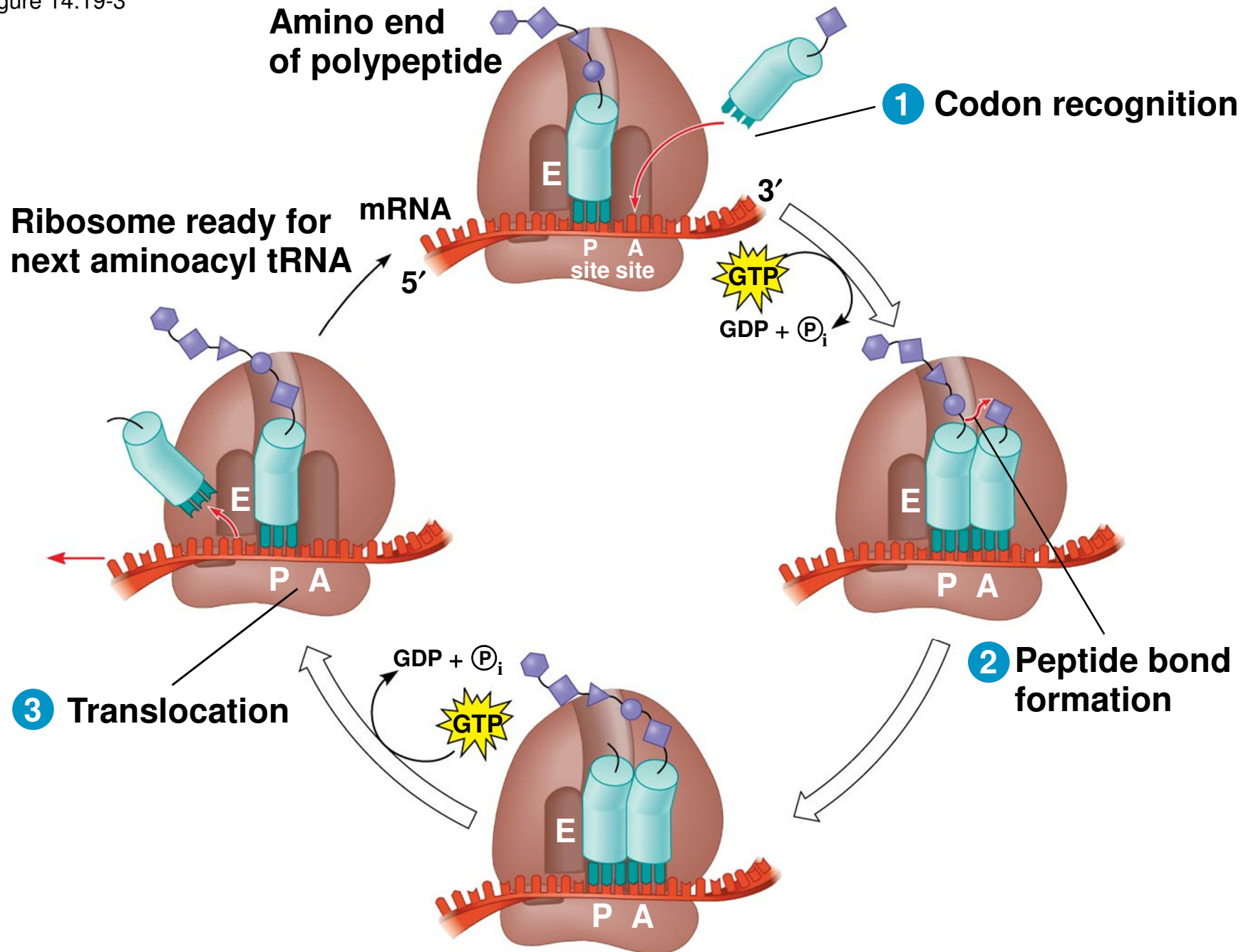


# *Elongation of the Polypeptide Chain*

---

- During elongation, amino acids are added one by one to the previous amino acid at the C-terminus of the growing chain
- Each addition involves proteins called *elongation factors* and occurs in three steps
  1. Codon recognition
  2. Peptide bond formation
  3. Translocation
- Translation proceeds along the mRNA in a 5' to 3' direction
  - Ribosome and mRNA move relative to each other

Figure 14.19-3

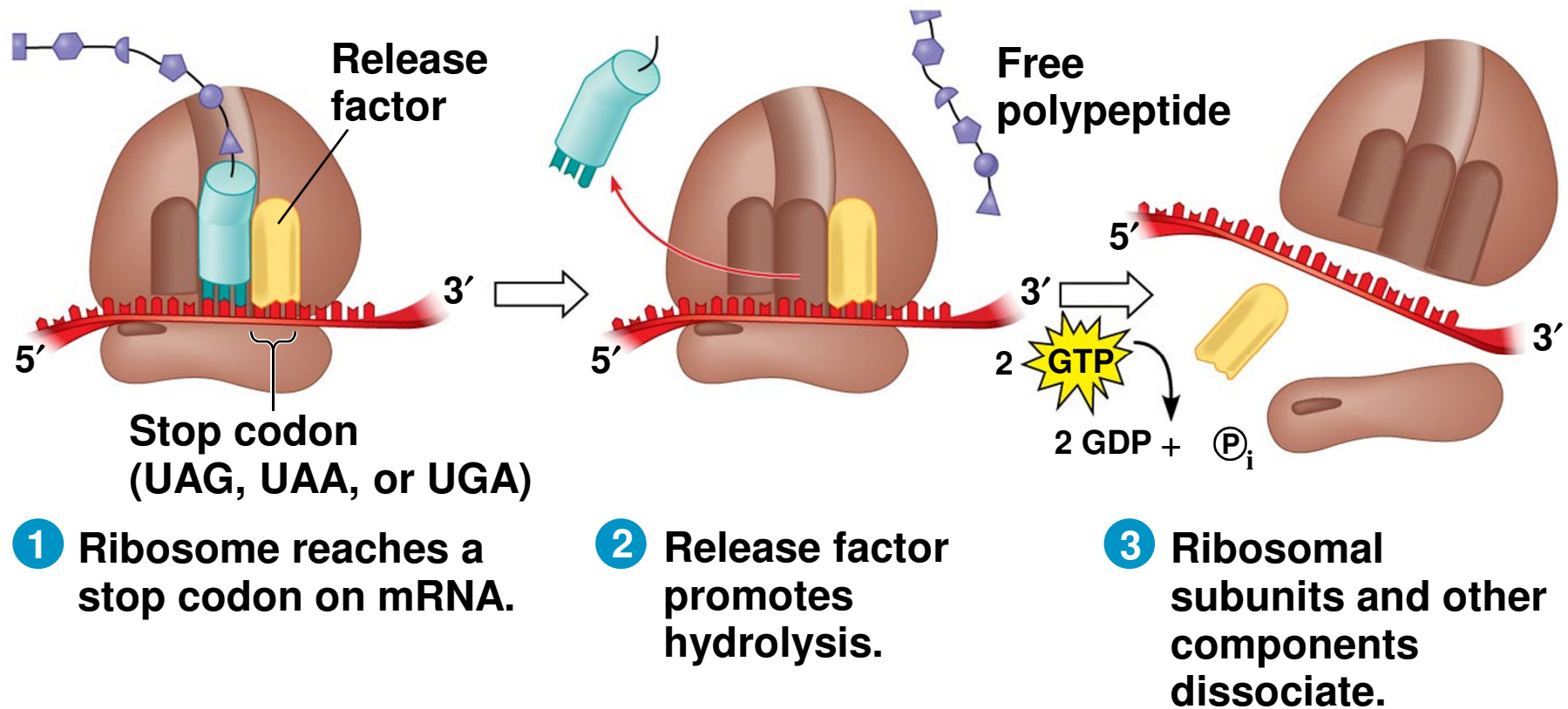


# *Termination of Translation*

---

- Termination occurs when a stop codon in the mRNA reaches the A site of the ribosome
- The A site accepts a protein called a *release factor*
- The release factor causes the addition of a water molecule instead of an amino acid
  - Breaks (hydrolyzes) the bond between the completed polypeptide and the tRNA in the P site
- This reaction releases the polypeptide through the exit tunnel
- The translation assembly then comes apart

Figure 14.20-3



# Completing and Targeting the Functional Protein

---

- Often translation is not sufficient to make a functional protein
- Polypeptide chains are modified after translation or targeted to specific sites in the cell

# *Protein Folding and Post-Translational Modifications*

---

- During synthesis, a polypeptide chain spontaneously coils and folds into its three-dimensional shape
  - A gene determines primary structure
  - Primary structure in turn determines shape
- A chaperone protein helps the polypeptide fold correctly
- Proteins may also require post-translational modifications before doing their jobs
  - Certain amino acids may be chemically modified by the attachment of sugars, lipids, or phosphate groups
  - Polypeptide chains may be cleaved or joined together

# *Targeting Polypeptides to Specific Locations*

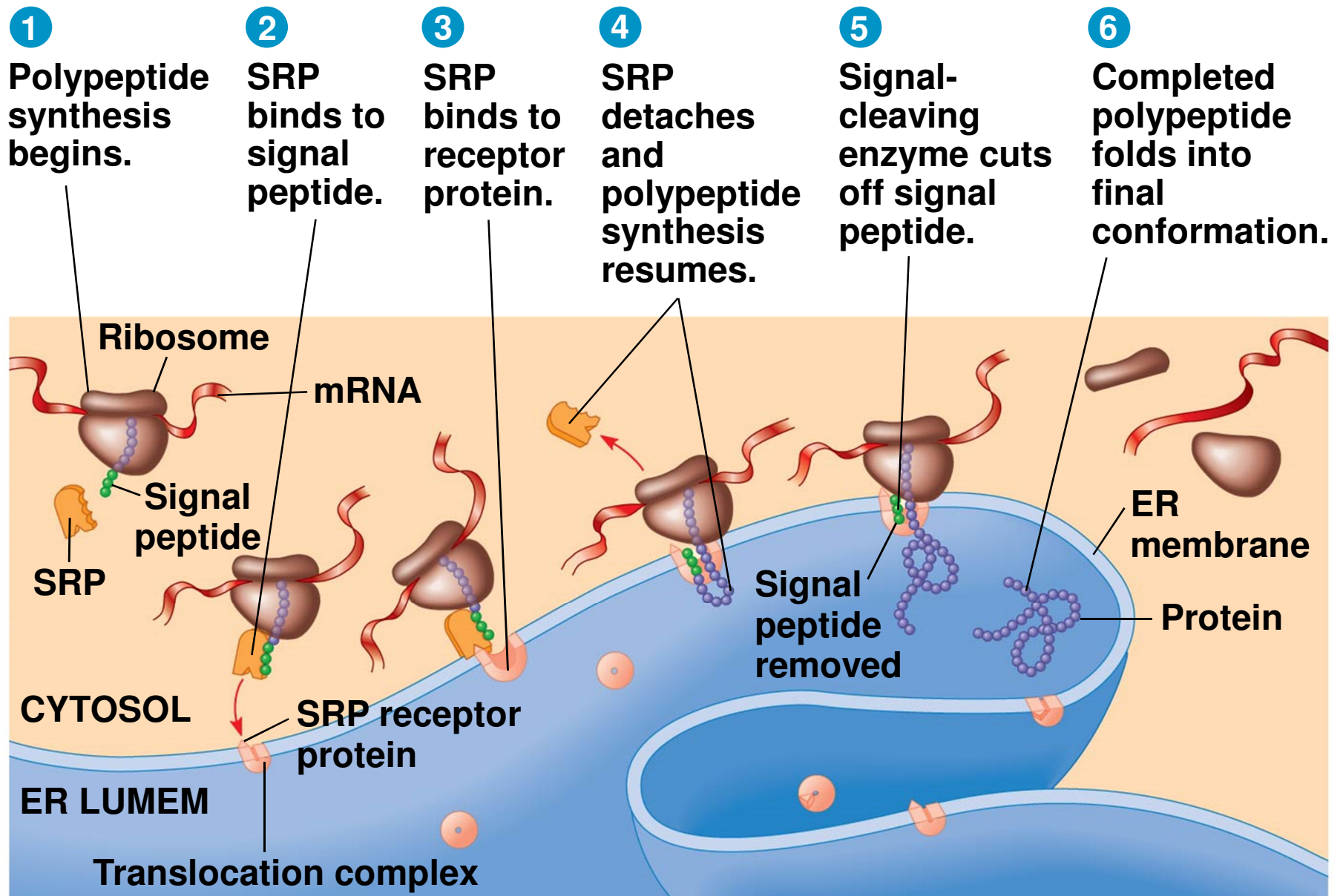
---

- Two populations of ribosomes are evident in cells:
  - Free ribosomes (in the cytosol)
  - Bound ribosomes (attached to the ER)
- Free ribosomes mostly synthesize proteins that function in the cytosol
- Bound ribosomes make proteins of the endomembrane system and proteins that are secreted from the cell



- 
- Polypeptide synthesis always begins in the cytosol
  - Synthesis finishes in the cytosol unless the polypeptide signals the ribosome to attach to the ER
  - Polypeptides destined for the ER or for secretion are marked by a **signal peptide**
  - A **signal-recognition particle (SRP)** binds to the signal peptide
    - The SRP brings the signal peptide and its ribosome to the ER

Figure 14.21

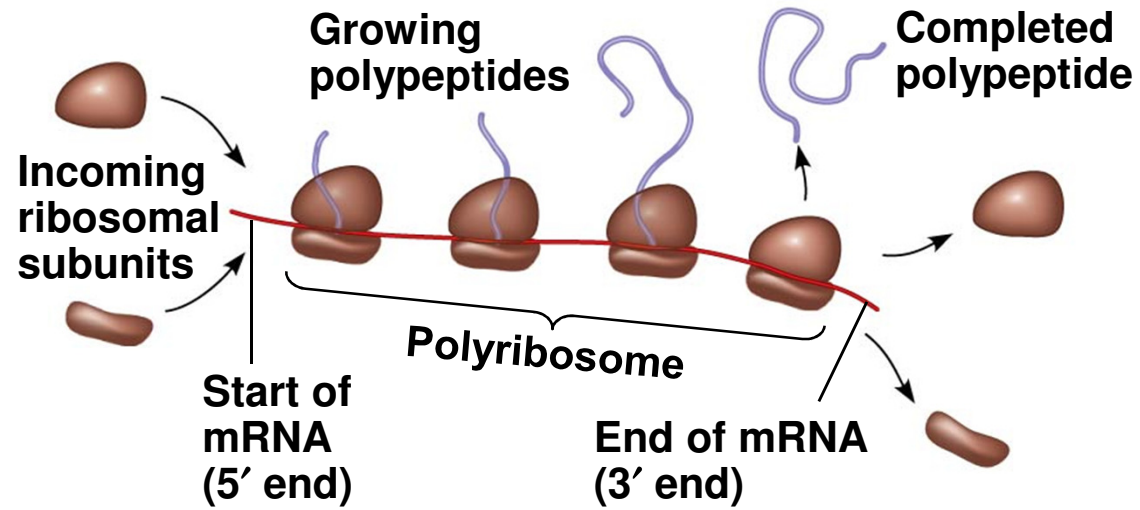


# Making Multiple Polypeptides in Bacteria and Eukaryotes

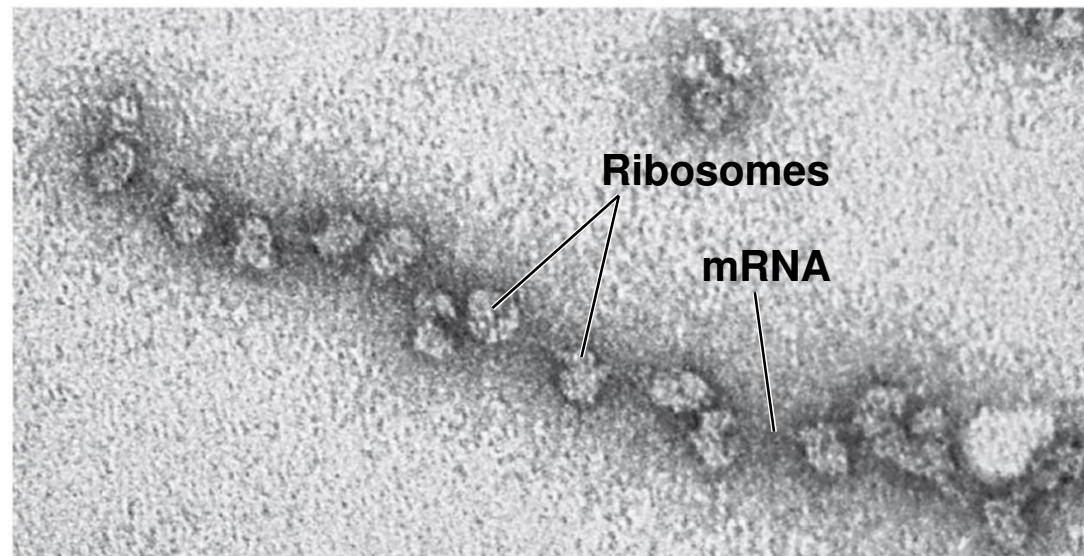
---

- In bacteria and eukaryotes multiple ribosomes translate an mRNA at the same time
- Once a ribosome is far enough past the start codon, another ribosome can attach to the mRNA
- Strings of ribosomes called polyribosomes (or polysomes) can be seen with an electron microscope
  - Enable a cell to make many copies of a polypeptide very quickly

Figure 14.22



(a) Several ribosomes simultaneously translating one mRNA molecule



(b) A large polyribosome in a bacterial cell (TEM)

- 
- Bacteria and eukaryotes can also transcribe multiple mRNAs from the same gene
  - In bacteria, the transcription and translation can take place simultaneously
  - In eukaryotes, the nuclear envelope separates transcription and translation
    - Provides a compartment for extensive RNA processing

Figure 14.23

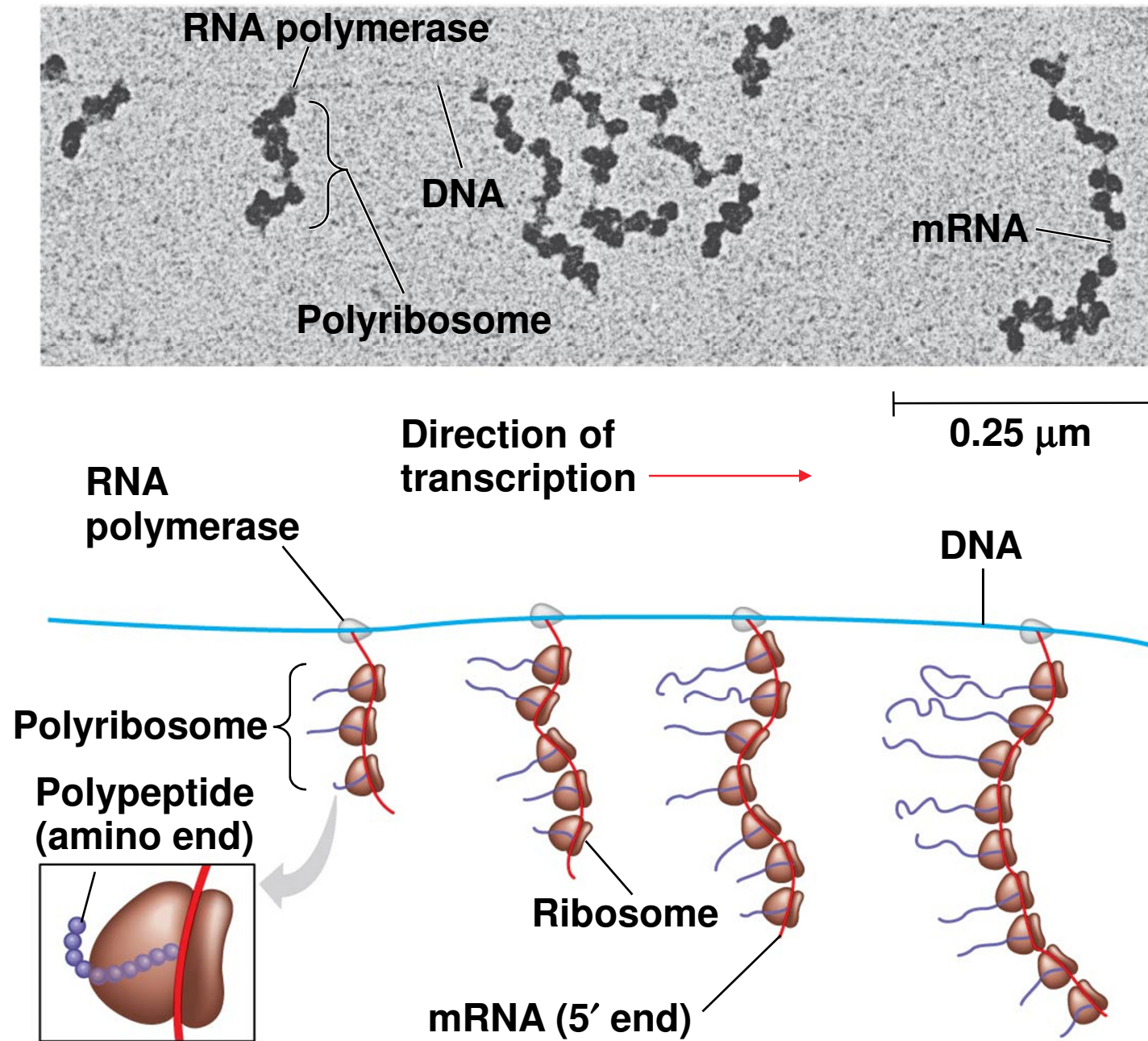
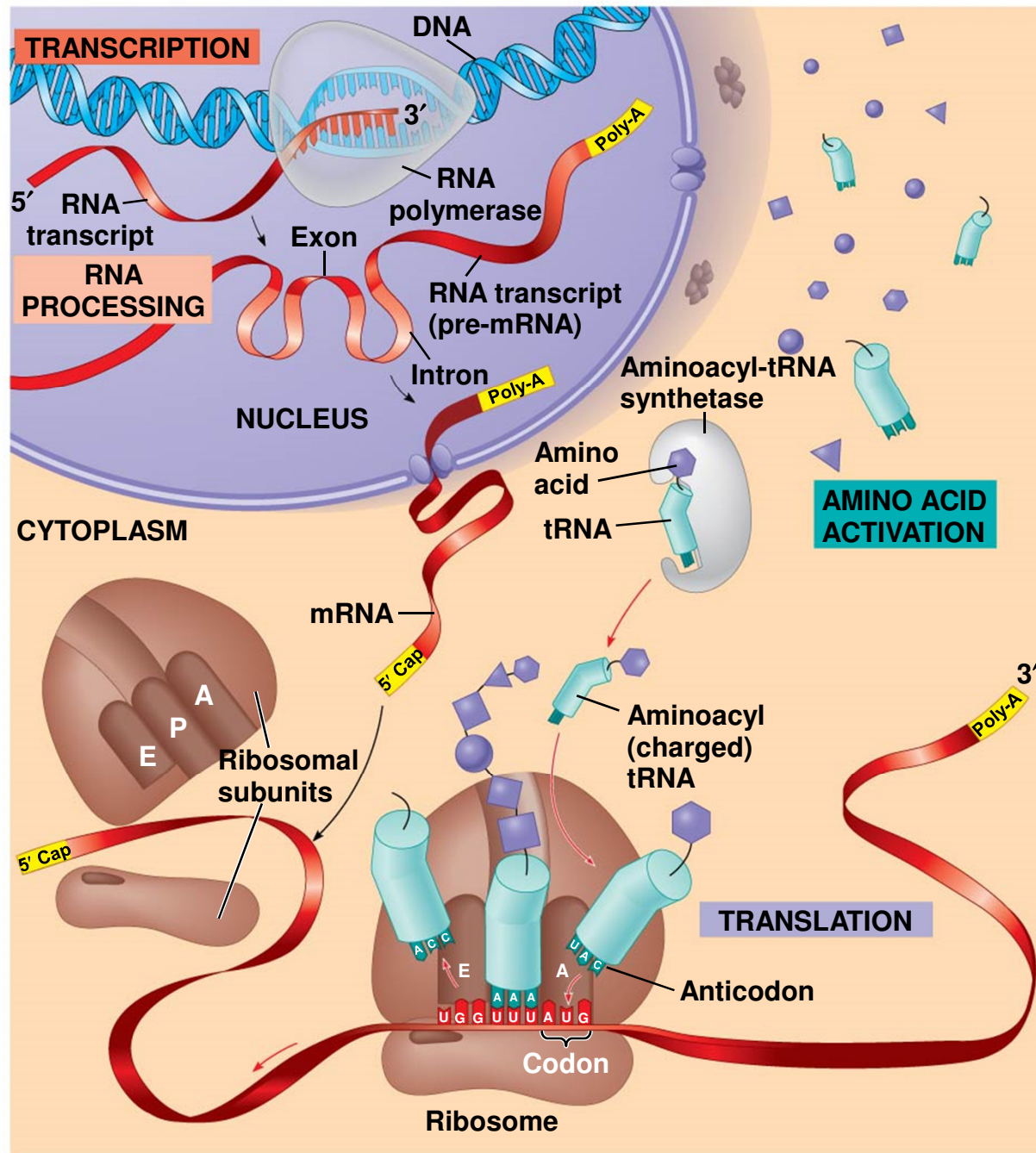




Figure 14.24



## Concept 14.5: Mutations of one or a few nucleotides can affect protein structure and function

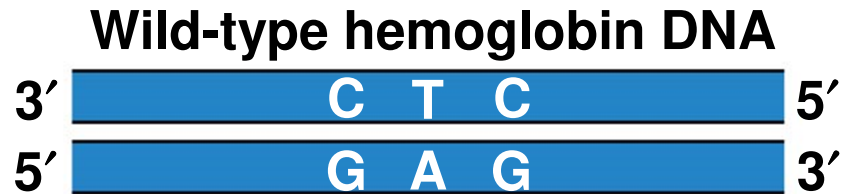
---

- **Mutations** are changes in the genetic material of a cell or virus
- **Point mutations** are chemical changes in just one or a few nucleotide pairs of a gene
- If the mutation occurs in a gamete or cell that gives rise to gametes, it may be transmitted to offspring
- The change of a single nucleotide in a DNA template strand can lead to the production of an abnormal protein

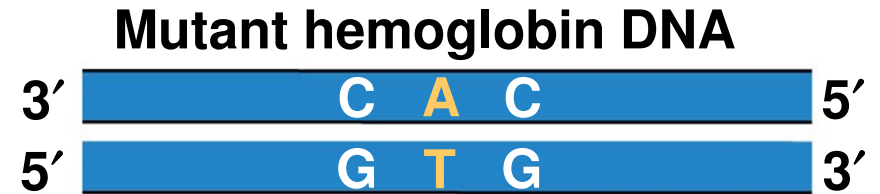


Figure 14.25

## Wild-type hemoglobin



## Sickle-cell hemoglobin



# Types of Small-Scale Mutations

---

- Point mutations within a gene can be divided into two general categories
  - Nucleotide-pair substitutions
  - One or more nucleotide-pair insertions or deletions

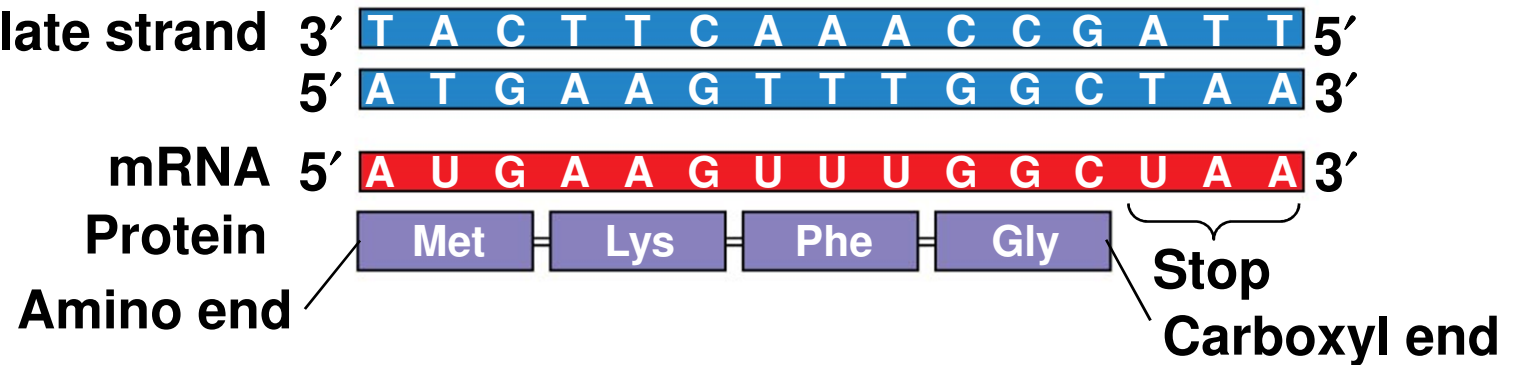
# *Substitutions*

---

- A **nucleotide-pair substitution** replaces one nucleotide and its partner with another pair of nucleotides
- **Silent mutations** have no effect on the amino acid produced by a codon because of redundancy in the genetic code
- **Missense mutations** still code for an amino acid, but not the correct amino acid
  - Substitution mutations are usually missense mutations
- **Nonsense mutations** change an amino acid codon into a stop codon
  - Nearly always lead to a nonfunctional protein

Figure 14.26a

## Wild type



## Nucleotide-pair substitution: silent

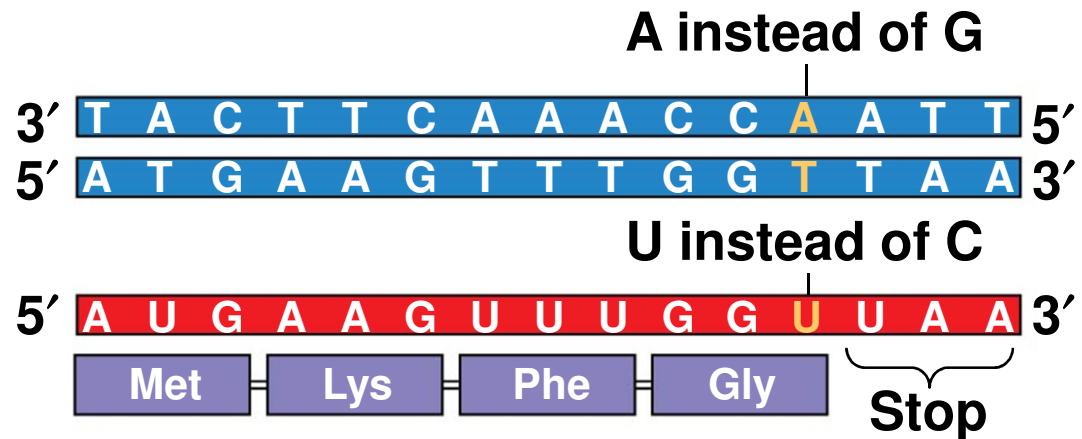
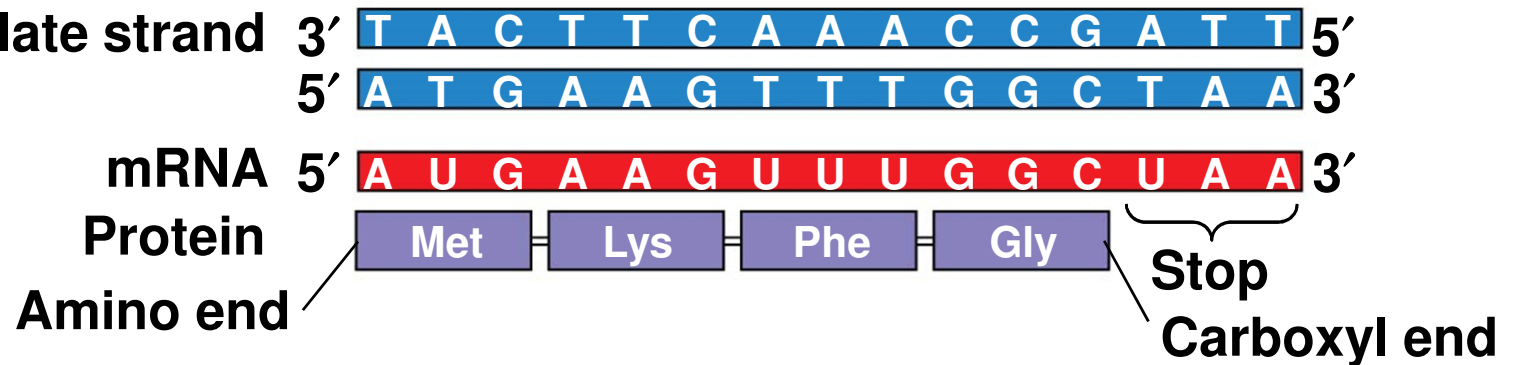


Figure 14.26b

## Wild type



## Nucleotide-pair substitution: missense

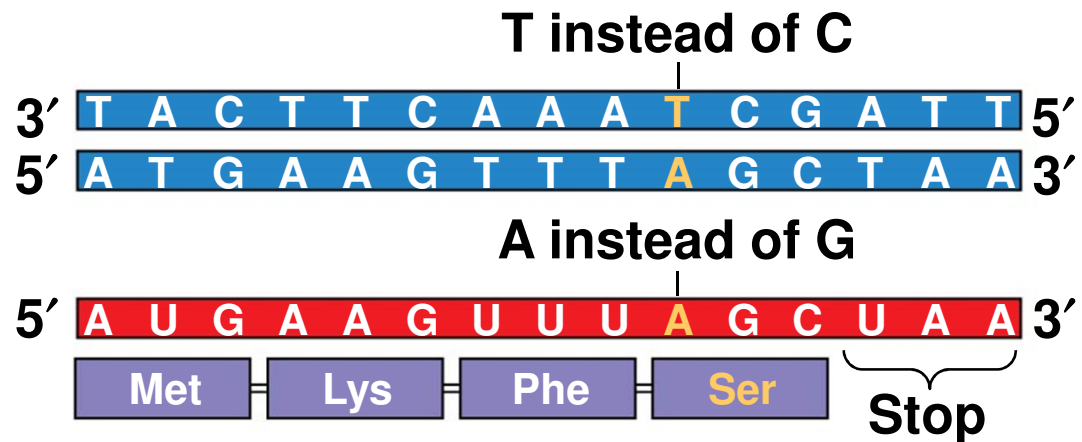
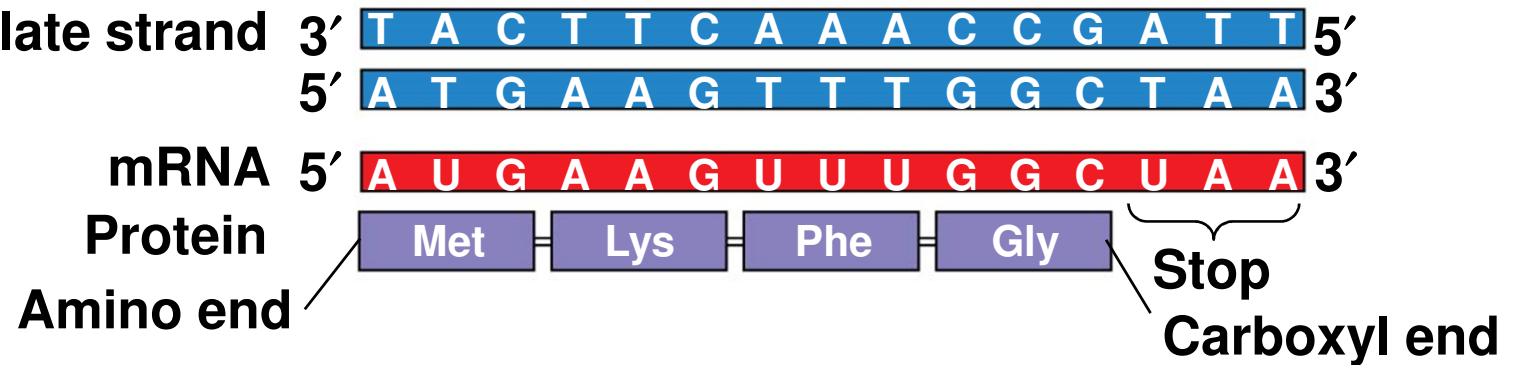
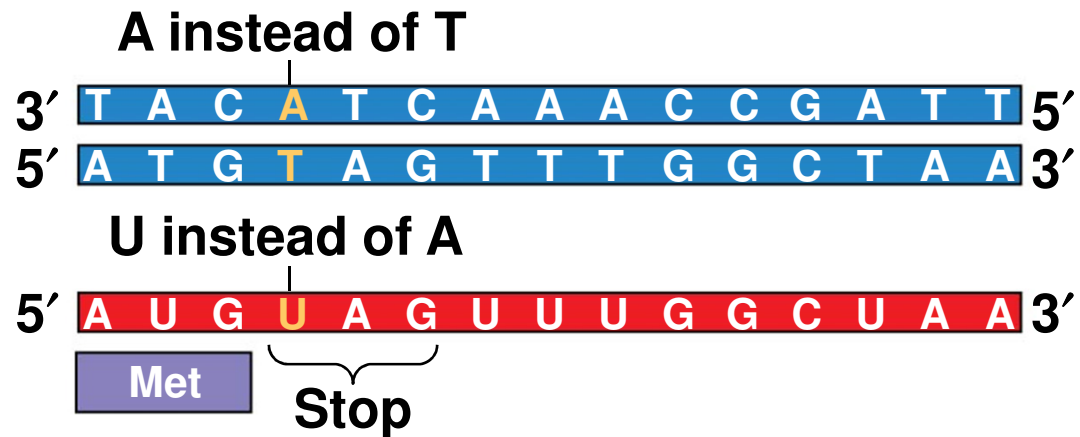


Figure 14.26c

## Wild type



## Nucleotide-pair substitution: nonsense



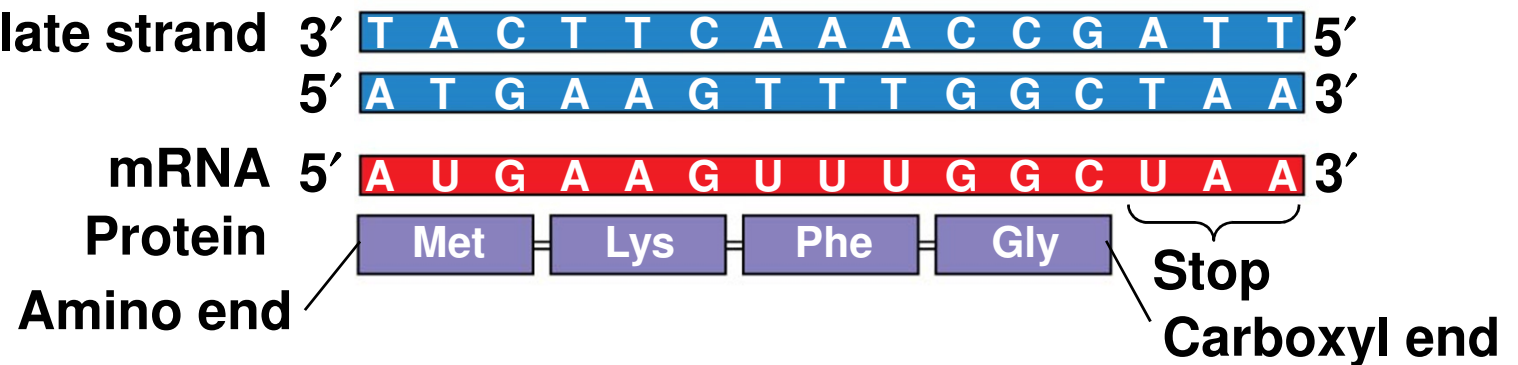
# ***Insertions and Deletions***

---

- **Insertions** and **deletions** are additions or losses of nucleotide pairs in a gene
  - These mutations have a disastrous effect on the resulting protein more often than substitutions do
- Insertion or deletion of nucleotides may alter the reading frame of the genetic message, producing a **frameshift mutation**
  - All of the nucleotides downstream of the deletion or insertion will be improperly grouped into codons
  - Results will be extensive missense
  - Resulting protein is typically nonfunctional
- Note: Inserting or deleting nucleotides in multiples of 3 will NOT shift reading frame
  - But will insert or delete a single amino acid

Figure 14.26d

## Wild type



## Nucleotide-pair insertion: frameshift causing immediate nonsense

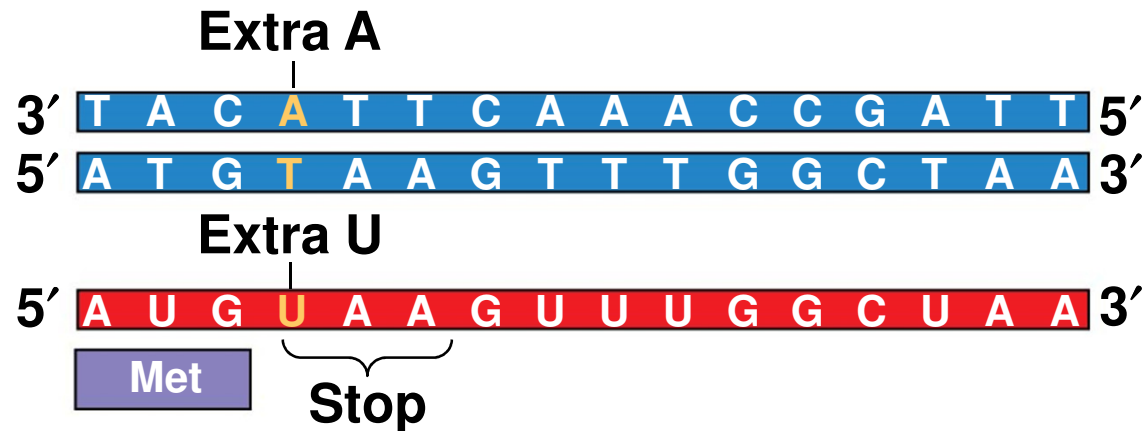
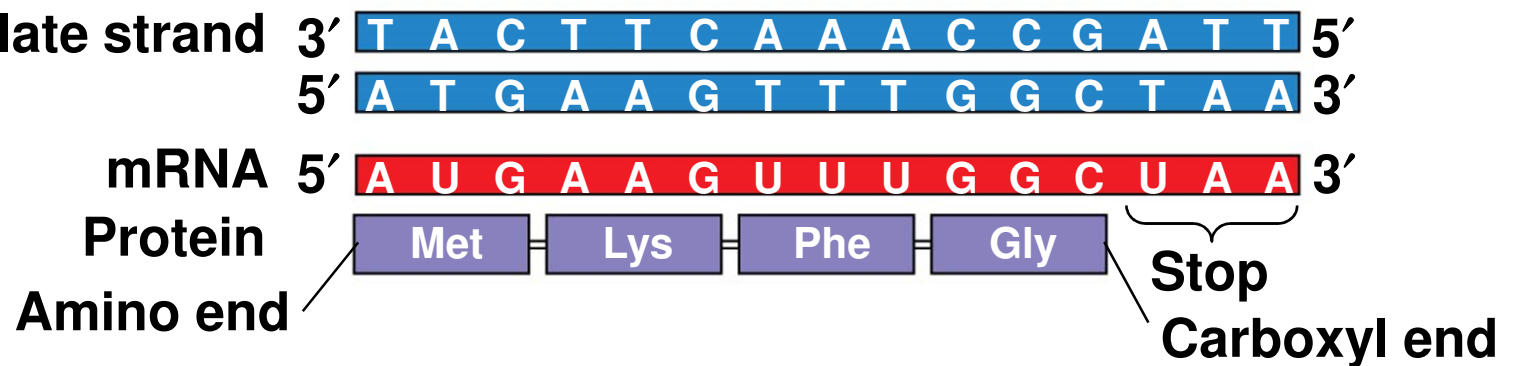




Figure 14.26e

## Wild type



## Nucleotide-pair deletion: frameshift causing extensive missense

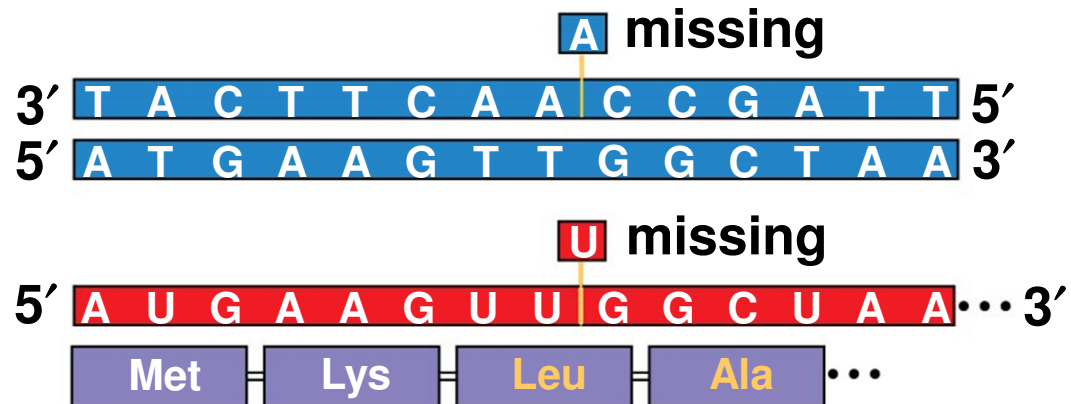
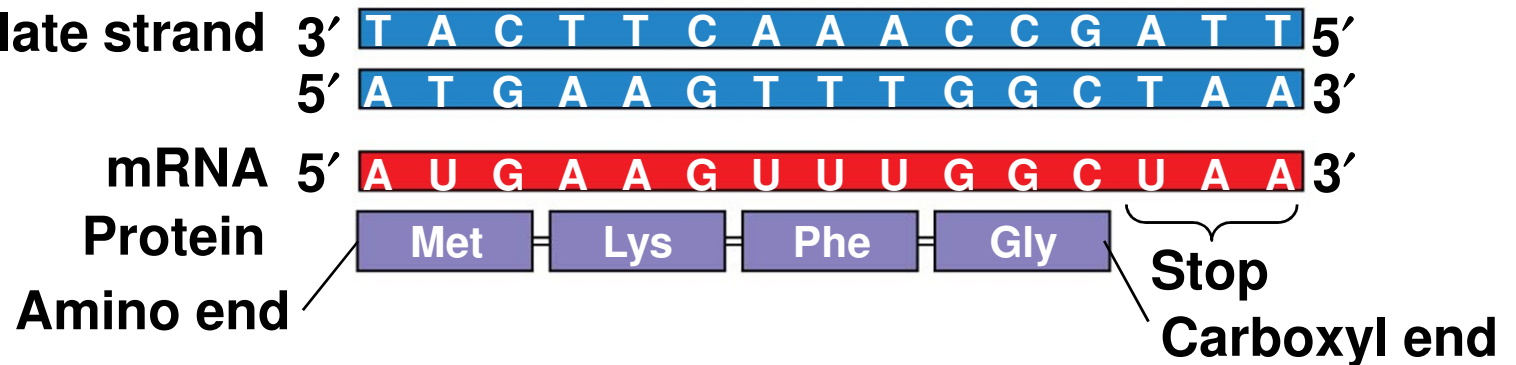
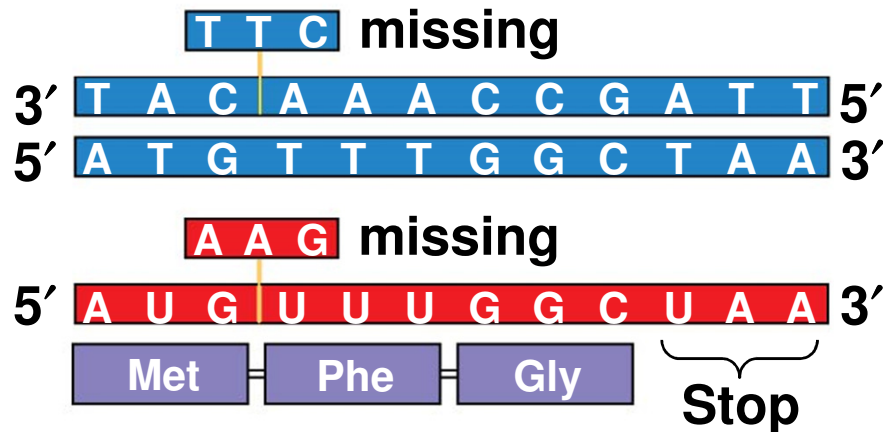


Figure 14.26f

## Wild type



**3 nucleotide-pair deletion: no frameshift, but one amino acid missing**



# Mutagens

---

- Mutations can occur during errors in DNA replication or recombination, leading to substitutions, insertions, or deletions
  - Often corrected by proofreading and repair systems
  - Otherwise, the incorrect base will be used as a template in future replications, resulting in a *spontaneous mutation*
- **Mutagens** are physical or chemical agents that can cause mutations
  - Most cancer-causing chemicals (carcinogens) are mutagenic, and the converse is also true
  - Radiation (X-rays and UV rays)

# What Is a Gene? *Revisiting the Question*

---

- The definition of a gene has evolved through the history of genetics
- We have considered a gene as
  - A discrete unit of inheritance that affects phenotype
    - Assigned to specific loci on chromosomes
  - A region of specific nucleotide sequence along the length of the DNA molecule of a chromosome
  - A DNA sequence that codes for a specific polypeptide chain

- 
- A gene can be defined as a region of DNA that can be expressed to produce a final functional product, either a polypeptide or an RNA molecule
  - Gene expression is precisely regulated